ANÁLISIS DISCRIMINANTE Y COMPARATIVO USANDO MÉTODOS KERNEL PARA IDENTIFICAR EL RENDIMIENTO ESCOLAR EN ESTUDIANTES DE OCTAVO GRADO DE LA INSTITUCIÓN EDUCATIVA TÉCNICA OCCIDENTE DE LA CIUDAD DE TULUÁ VALLE DEL CAUCA.

JOSÉ ORLANDO RODRÍGUEZ RODRÍGUEZ

UNIVERSIDAD TECNOLOGÍCA DE PEREIRA
FACULTAD DE CIENCIAS BÁSICAS.
MAESTRÍA EN ENSEÑANZA DE LA MATEMÁTICA.
PEREIRA
2018

ANÁLISIS DISCRIMINANTE Y COMPARATIVO USANDO MÉTODOS KERNEL PARA IDENTIFICAR EL RENDIMIENTO ESCOLAR EN ESTUDIANTES DE OCTAVO GRADO DE LA INSTITUCIÓN EDUCATIVA TÉCNICA OCCIDENTE DE LA CIUDAD DE TULUÁ VALLE DEL CAUCA.

JOSÉ ORLANDO RODRÍGUEZ RODRÍGUEZ

Trabajo de Grado para optar al título de

Magister en enseñanza de la matemática.

Director

Dr. EDGAR ALIRIO VALENCIA A.

UNIVERSIDAD TECNOLOGÍCA DE PEREIRA

FACULTAD DE CIENCIAS BÁSICAS.

MAESTRÍA EN ENSEÑANZA DE LA MATEMÁTICA.

PEREIRA

2018

DEDICATORIA

Dedico este trabajo primero que todo a Dios porque día a día se convierte en mi fortaleza para seguir a delante, no bajar los brazos ni creer que puedo dar algo por perdido; además, me brindó la energía, saberes y conocimientos para terminar con gusto este proyecto.

A mi esposa y mis hijos por el constante apoyo y significado moral y mental. A mi padre Arturo (Q.E.P.D.) que está en el cielo y que siempre estuvo vigilante esperando cosas gratas de mí. A mi madre por todo ese apoyo tan grande que ha tenido en todo mi proceso educativo y aun a mi edad siempre mantiene pendiente.

AGRADECIMIENTOS

A mi director de tesis que me acogió con mucho agrado después de haber pasado por muchas dificultades. A todos y cada uno de los docentes de la maestría que pusieron en mi todos sus conocimientos y que me permitieron la formación necesaria para llegar a construir este trabajo de grado que es tan valioso para mi vida. A toda la formación y responsabilidad que me brindaron mis padres. A José Rodrigo González director del programa por permitir la finalización de la tesis.

TABLA DE CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

LISTA DE ANEXOS

RESUMEN

El análisis se realiza a 5 grupos de estudiantes de un mismo grado, en este caso grado octavo con 5 cursos desde 8-1 hasta 8-5, se eligen algunas variables o características que distinguen a los estudiantes de buen o bajo rendimiento académico. Se evalúan quince variables de las cuales hay una variables dependiente o categórica llamada variable de agrupación nota promedio (nivel de competencia) y catorce (14) variables independientes o predictoras organizadas como sigue: dos de responsabilidad, dos de acompañamiento, dos de motivación, dos de expectativa, una de convivencia, dos de nivel familiar, dos de cumplimiento de normas, otra de rendimiento académico. La muestra es de 184 individuos, la cual se divide en casos de 4, 3 y 2 grupos; además, la nota límite que identifica el nivel Superior del nivel Bajo es la calificación de 6,0 a 10,0 puntos indicando buen rendimiento y tipificado como Superior, Alto y Básico en el caso de 4 grupos, para el caso de 3 grupos Medio y Alto, para el caso de 2 grupos o último análisis se tipifica como Aprueba, aplicando a cada caso toda la estructura discriminante y comparativamente con el método kernel.

Las variables elegidas en el análisis discriminante fueron: Perdidas (número de asignaturas perdidas), Padrespend (sus padres están al pendiente del desarrollo de sus actividades académicas) y Cartracasa (le ponen excesiva carga de trabajo en su casa) y en el método kernel Perdidas, Padrespend, Cartracasa y Entregatareas (entrega a tiempo sus tareas), siendo muy similares los dos métodos en la determinación de variables del rendimiento académico de los estudiantes de grado octavo; además, dos fueron los grupos finalmente obtenidos (Bajo y Aprueba) a los que se redujo el análisis discriminante y el kernel comparativamente buscando identificar la mejor clasificación de individuos, reducir variables y grupos para una mejor y efectiva discriminación; además, se halla el porcentaje de correcta clasificación dada en los grupos con todas las variables en los dos métodos y en el análisis discriminante se complementa un último análisis con las variable que muestran mayor poder discriminante y una clasificación de individuos de más del 80% que al final permiten obtener la función discriminante, la cual se encarga de clasificar individuos y la función apriori que clasifica nuevos individuos.

Para el método kernel se aplica la función Gaussiana con diferentes estimaciones y ensayos de varianza, desde 0.01 hasta 1.0, para determinar cuál es la más eficiente para clasificación de individuos y selección de variables con mayor poder discriminante, similar al análisis discriminante, teniendo en cuenta que para la aplicación de la función kernel también se hace necesario la medias de grupo, las medias muestrales y las desviaciones estándar en cada variable según el ensayo.

Todo lo anterior logra mostrar que, el método kernel clasifica mejor los individuos, especialmente, para 4 y 3 grupos; además, la varianza muestral y las varianzas de ensayo 0.50 y 0.70 resultaron ser las más efectivas para clasificar los individuos con este método.

PALABRAS CLAVES: Análisis discriminante, método kernel, rendimiento académico, variable dependiente o categórica, variable independiente o predictoras.

OBJETIVOS

Objetivo general

- Realizar análisis discriminante comparativo para el rendimiento académico general en estudiantes de octavo grado de la institución educativa técnica occidente.

Objetivos específicos

- Establecer u ordenar nuevos individuos dentro de grupos previamente reconocidos o definidos analizando si existen diferencias en cuanto a su rendimiento.
- Elaborar procedimientos de clasificación sistemática de individuos de origen conocido.
- Obtener una función capaz de clasificar en varios grupos y a crear una función capaz de distinguir con la mayor precisión posible a los miembros de cada grupo.
- Proponer el análisis discriminante descriptivo y un método kernel como comparativos que permitan apoyar el estudio del rendimiento académico de los estudiantes de octavo grado de la institución.
- Considerar y estudiar que variables se pueden tomar en cuenta para aplicar los métodos de análisis discriminante y el método kernel para encontrar los factores que influyen en el rendimiento académico de los estudiantes de grado octavo.

INTRODUCCIÓN

La gran mayoría de los estudiantes de bachillerato presentan ciertas situaciones que les puede impedir llevar un excelente rendimiento académico; por esto, el promedio y nivel académico o rendimiento va ligado a variables que se pueden clasificar y discriminar según su importancia para un nivel educativo bueno y de calidad en la institución educativa técnica occidente de Tuluá.

Según Bloom (1976), el rendimiento académico como resultado consiste en "las diferentes formas que se emplean para cada etapa o nivel de aprendizaje escolar, además, son la base para predecir si se pueden pasar a la siguiente etapa". Esto permite decir que las distintas variables son de vital importancia.

Por otro lado el rendimiento estudiantil no solo es una calificación y viene concatenada en una política de admisión (Bonucci, 1997), lo que conlleva a afirmar que son muchas las variables inherentes en un proceso de aprendizaje que en la medida que sean identificadas con un aceptable índice de precisión pueden ser manejadas y controladas para mejorarlas. El rendimiento académico no es el producto de una sola capacidad sino que es el resultado de varios factores que interactúan en la persona que aprende (Gómez et al, 2011).

Al tener en cuenta la medición del rendimiento académico en diferentes apuntes, también se consideran los factores que influyen en este, el presente estudio toma en cuenta algunos antecedentes importantes:

Análisis del rendimiento académico mediante un modelo logit" (según Ibarra y Michalus, 2010). En este trabajo se analiza el rendimiento académico de los estudiantes de la Facultad de Ingeniería de la Universidad Nacional de Misiones y define al rendimiento académico como el promedio de materias aprobadas anualmente, y mediante la técnica estadística multivariada de regresión logística se determina la incidencia de factores de índole personal, socioeconómica y académica. Los resultados obtenidos permiten concluir que las variables significativas del rendimiento académico son: el promedio de calificaciones del nivel medio, el tipo de Institución donde cursó estos estudios y el número de asignaturas aprobadas en el primer año de carrera,

siendo este último factor el más relevante, destacando la importancia de esta primera etapa de la carrera en los posteriores resultados académicos del estudiante.

Rodríguez Ayán María N. (2007) Análisis multivariado del desempeño académico de estudiantes universitarios de Química (Tesis doctoral). El propósito de esta investigación fue definir un indicador del desempeño académico estudiantil basado en los créditos académicos, como alternativa al indicador tradicional basado en el promedio de calificaciones. Se estudia comparativamente el comportamiento de ambos indicadores mediante modelado estadístico multivariante, a partir de variables explicativas sociodemográficas, académicas y motivacionales.

Se construyen y comparan modelos de regresión lineal y logística y modelos de ecuaciones estructurales, en dos grupos independientes extraídos de una misma población. Los resultados sugieren que el comportamiento de ambos indicadores es similar. En cuanto a los modelos de regresión, se confirma la pérdida de potencia para detectar efectos significativos al categorizar la variable criterio para su modelado mediante regresión logística. Respecto a los modelos de ecuaciones estructurales, los modelos de variables latentes que utilizan agrupaciones de ítems resultan una alternativa atractiva frente a los modelos de rutas.

Marquín T. María J. (2017) Predicción del rendimiento académico mediante técnicas del análisis multivariado en la asignatura de algebra lineal. Universidad Tecnológica de Pereira (UTP). El presente estudio describe cómo se aplicó la técnica de regresión logística y análisis discriminante para encontrar las variables significativas que inciden en el rendimiento académico de la asignatura de Algebra Lineal. Se corrieron 11 modelos en el programa SPSS utilizando la técnica de regresión logística de los cuáles solo seis (modelos 3, 7, 8, 9, 10 y 11) inciden en el rendimiento académico de la asignatura, siendo la variable rendimiento académico de matemáticas durante el bachillerato la que aporta mayor porcentaje de clasificación. Se utilizó la técnica de análisis discriminante con el objetivo de probar si el porcentaje de clasificación de los tres modelos con mayor significancia sobre la variable dependiente (modelos 3, 10 y 11) mejora con este método, obteniendo como resultado que el modelo 3 sigue siendo el modelo con mayor porcentaje de global de clasificación (90,8%)

Saavedra Lutzgardo, Ramos Julio C., Mitacc Máximo C., Del Águila Víctor R. (2017). Artículo "Detección temprana del rendimiento académico de estudiantes universitarios de primer ciclo mediante el análisis discriminante". En este estudio se ha procurado identificar a los ingresantes que aprobarían a lo más dos de los cinco cursos en los que se matricularon para el

segundo semestre 2016 del Programa de Estudios Generales de la Universidad de Lima. Dicha identificación se basó en modelos de predicción, construidos con datos del semestre 2016-1 mediante el uso de análisis discriminante. La población de ingresantes se dividió en tres dominios de estudio y se construyeron modelos independientes de predicción para el rendimiento académico utilizando las funciones de clasificación de Fisher, evaluadas mediante los indicadores de rendimiento y la curva Receiver Operating Characteristic (ROC).

Mendoza Adel y Herrera Roberto. (Barranquilla 2013). Predicción del rendimiento académico de los estudiantes de la universidad del atlántico, basado en la aplicación del análisis discriminante. Este artículo propone que la Universidad del Atlántico implemente el uso del análisis discriminante, que es un modelo estadístico multivariado que tiene como objetivo encontrar la combinación lineal de las variables independientes que mejor permite diferenciar (discriminar) a los grupos. Una vez encontrada esa combinación (la función discriminante) podrá ser utilizada para clasificar nuevos casos; para esto utiliza toda la estructura discriminante.

Todos los trabajos anteriores muestran lo interesante del uso del análisis discriminante, pero también para los métodos kernel se han hecho algunas aplicaciones como son:

Eddy Rodríguez, Carlos Vinante, Marihebert Leal. (2009). Universidad de Oriente Venezuela. Enfoque óptimo del método kernel cuadrados mínimos parciales. En este trabajo se presentar un enfoque óptimo de los algoritmos del método Kernel de Cuadrados Mínimos Parciales; además, para la investigación se usó una metodología descriptiva-comparativa partiendo de la consulta en fuentes bibliográficas, páginas web y la comparación computacional de los algoritmos de estos métodos mediante el estudio de ejemplos con características específicas, dando como resultado que (KPLS) el método kernel de Cuadrados Mínimos Parciales con descomposición en valores singulares (SVD) presenta tiempos de entrenamientos cortos y un diseño estructural sencillo. De esta investigación se concluye, que el método KPLS con SVD es una excelente opción para el modelado de problemas de regresión no lineal.

González Nelson H. (2013) Universidad Nacional De Colombia. Métodos de Kernels en secuencias para la clasificación de residuos catalíticos en sitios activos de enzimas. Este trabajo presenta una metodología de solución al problema de clasificación de residuos catalíticos en sitios activos de enzimas. Esta metodología está basada en el aprendizaje de máquina específicamente en las máquinas de soporte vectorial (MSV); que junto a las funciones kernel permite clasificar residuos en enzimas a partir de su secuencia. En la metodología planteada, en

primer lugar encontramos la información biológica de los residuos integrada con la representación en secuencia de la enzima que lo contiene; esto por medio de las funciones kernel gaussiano y string, respectivamente. Posteriormente; el algoritmo jerárquico es aplicado para obtener un número de grupos inicial para el algoritmo de agrupación k-medias; obteniendo como resultado cinco grupos de enzimas. Por último, para cada grupo se desarrolló un sistema basado en MSV. La estimación del error de generalización después de validación cruzada es usada como criterio de desempeño del modelo.

Velasco P. Víctor. (2016). Departamento de Ingeniería Informática. Universidad Autónoma de Madrid. Aprendizaje multitarea mediante procesos gaussianos para clasificación (Kernel). En este trabajo se muestra que el aprendizaje multitarea es una técnica de aprendizaje automático que consiste en entrenar un algoritmo de aprendizaje automático para aprender múltiples tareas usando una representación compartida para todas ellas. Aunque el enfoque tradicional en el campo del aprendizaje automático consiste en aprender una única tarea, en ciertos problemas aprender varias tareas relacionadas al mismo tiempo aumenta la capacidad predictiva. Con el trabajo se propone comprobar la utilidad del aprendizaje multitarea a la hora de afrontar problemas de clasificación. Para ello se utiliza los modelos conocidos como procesos gaussianos, que recientemente han tenido una gran acogida en la comunidad de investigadores en el campo del aprendizaje automático. Para reconocer patrones, los procesos gaussianos necesitan de una función núcleo (o kernel) suficientemente expresivo. En este trabajo utilizamos un kernel conocido como Spectral Mixture Kernels. Estos kernels tienen una propiedad que los hacen especialmente útiles: podemos aproximar cualquier función de covarianza estacionaria con una precisión arbitraria.

Estos trabajos expuestos permiten conocer la utilidad de los métodos aplicados en el presente trabajo, además, identificar con que elementos se puede abordar el estudio que se realiza; de este análisis de situaciones o antecedentes se puede definir que el promedio de todas las áreas, acompañado de otras variables de índole social y cultural pueden convertirse en un buen elemento de medición para identificar el rendimiento académico y además permita, implementar políticas tendientes a mejorarlo; conviene además, conocer las causas de las dificultades presentadas por los estudiantes.

Con este análisis discriminante y método kernel se buscan resultados que posteriormente ayudarán a otras instituciones en estudios de rendimiento académico, ya sea un área específica o

en todo el pensum académico, como también, aspectos relacionados con la influencia del entorno. El conocimiento de las variables discriminantes nos lleva a determinar algunas de las dificultades concretas que se encuentran en los estudiantes.

El estudio que se realiza en esta tesis va dirigido a la institución educativa técnica occidente de Tuluá, específicamente a grado octavo en diferentes cursos cuyo objetivo es encontrar las variables que intervienen y pueden contribuir al rendimiento académico de los estudiantes, para esto se tuvieron en cuenta 184 individuos encuestados y agrupados en 5 cursos de grado octavo.

El trabajo en general contiene un marco teórico, seguido por la metodología y el análisis de los resultados de cada uno de los métodos aplicados. En la primera parte se abordan conceptos de rendimiento académico, análisis discriminante, método kernel y definiciones, además, teóricos de diferentes autores sobre el rendimiento académico.

En la metodología se describen algunos aspectos de la población y las variables utilizadas en el estudio, interpretación del análisis discriminante, supuestos, estadísticos, seguido por los resultados y análisis del método; se hace el mismo proceso con el método kernel llegando a finalizar con los resultados y análisis.

El método del análisis discriminante se utiliza para verificar los porcentajes de clasificación, las variables de mayor poder discriminantes y la función discriminante haciendo uso de los estadísticos Lambda de Wilks, Menor Razón F, Chi-cuadrada, Anova, correlación canónica, matrices, funciones, etc. Por otro lado, el método kernel se utiliza para clasificar individuos y hallar variables discriminantes utilizando las medias, desviación estándar, varianzas estimadas y de ensayo.

1. MARCO TEÓRICO

1.1. El rendimiento académico

Es mucho lo que podemos decir cuando de rendimiento escolar se trata, muchas son las variables a tratar que se condensan en un aspecto fundamental, el interés del estudiante y con esta su eficiencia académica. El bajo rendimiento académico según, Serrano y Adell, (2002) se produce por problemas muy específicos tales como: rechazo en la escuela, poco interés académico, deterioro en las relaciones interpersonales con los padres u otros estudiantes y aun con el mismo docente; además, las tecnologías han aumentado y han influenciado las malas prácticas de algunos estudiantes en el aula, como también los aspectos de índole social y cultural.

Martínez Hernández y Núñez (1987) expresa que: "El bajo rendimiento académico se debe a múltiples factores como falta de motivación y comunicación de los padres de familia hacia el estudiante, considerando que estos problemas son similares en todas partes; además, es necesario crear estrategias como: visitas domiciliarias a los padres de familias, estudiantes monitores, mejoramiento del currículo escolar en el aula de clase o en la casa; además, es importante que los docentes realicen sus clases recreativas grupales para animar a los estudiantes a relacionarse entre compañeros".

Se puede suponer que el niño que tiene dificultades en el aprendizaje muestre un bajo rendimiento escolar y al mismo tiempo un conflicto de personalidad que no puede expresarse con palabras. Estudiar, realizar tareas escolares, acreditar un curso, implican trabajo; eso lo sabemos todos, y que el trabajo es un gasto de energía.

Si los niños no invierten cierta cantidad de energía en las labores escolares, sería necesario preguntarnos por qué no lo hacen. Tal vez ahí encontremos que el niño necesita de la motivación y la atención de sus padres y docentes, pues estos factores son el alimento para el deseo y las ganas de aprender, ya que encontrar a un niño apático y sin interés de participar y trabajar en las

actividades dentro y fuera del aula puede deberse a que no se le ha brindado la atención requerida para sentirse comprendido y comprometido para sobresalir dentro del ámbito escolar.

Para tratar y analizar este tema desde un aspecto numérico hemos recurrido a los fundamentos de la estadística inferencial actual[1], en donde muchos de los métodos utilizados y que le dan veracidad a los estudios estadísticos son debidos a **Ronald Aylmer Fisher** y su metodología estadística tal como hoy la conocemos y vemos que la aplicación de estos métodos estadísticos a áreas tan diversas, nos lleva a la búsqueda de respuestas a interrogantes planteados por la ciencia y que impulsan el desarrollo de nuevos métodos estadísticos como el análisis discriminante, pero también se pretende identificar y utilizar otro método como comparativo para la clasificación como lo es el método kernel con el uso de sus fórmulas representativas de funciones y que tiene mucha utilidad en la estadística actual. Los métodos kernel[2] son una familia de algoritmos cuyo elemento común y pieza fundamental en todo es la función kernel, su utilidad en el análisis de datos reside en la representación de la información. Este tipo de métodos presentan la ventaja de que son aplicables a cualquier tipo de datos, además se pueden aplicar algoritmos lineales y obtener con ellos soluciones no lineales. Buscar e identificar un método de clasificación de individuos y variables del rendimiento escolar es una tarea ardua y el propósito de este estudio es ayudar en esta tarea.

1.2. El análisis discriminante

- Para este análisis se requiere de una variable categórica y el resto de variables son de intervalo o de razón y son independientes respecto de ella; significando esto que son variables numéricas cuyos valores representan magnitudes y la distancia entre los números de su escala es igual. Con este tipo de variables podemos realizar comparaciones de igualdad/desigualdad, establecer un orden dentro de sus valores y medir la distancia existente entre cada valor de la escala; además, las variables de razón poseen las mismas características de las variables de intervalo, con la diferencia que cuentan con un cero absoluto; es decir, el valor cero (0)

1 http://www.monografias.com/trabajos94/rendimiento-academico-nicaragua/rendimiento-academico-nicaragua.shtml#ixzz56muo7b10

2 López D. Ana (2017). Fundamentos Matemáticos de los Métodos Kernel para Aprendizaje Supervisado. Facultad de matemáticas. Universidad de Sevilla. España. Pag. 40

- Representa la ausencia total de medida, por lo que se puede realizar cualquier operación aritmética[3].

- Es necesario que existan al menos dos grupos, y para cada grupo se necesitan dos o más casos.

- El número de variables discriminantes debe ser menor que el número de objetos menos 2 $x_1,\ldots,x_p$, donde $p < (n-2)$ y n es el número de objetos. Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.

- El número máximo de funciones discriminantes es igual al mínimo entre el número de variables y el número de grupos menos 1, $min(g-1,\ p)$, con g grupos, $p < (n-2)$ variables.

- Las matrices de covarianzas dentro de cada grupo deben ser aproximadamente iguales.

Las variables continuas deben seguir una distribución normal multivariante[4]. Durante un estudio, a menudo hay preguntas que afectan al investigador y que deben ser respondidas. Estas preguntas incluyen situaciones como ¿son diferentes los grupos?, ¿En qué variables son más discrepantes los grupos? ¿Uno puede predecir a qué grupo pertenece una persona usando tales variables? etc. Para responder a estas preguntas y muchas más, el **análisis discriminante** es bastante útil[5]; teniendo en cuenta que es una técnica que utiliza el investigador para analizar los datos de estudio cuando el criterio o la variable dependiente es categórico y el predictor o la variable independiente es de intervalo en la naturaleza. El término variable categórica significa que la variable dependiente se divide en varias categorías; además, se desarrollan funciones discriminantes que no son más que la combinación lineal de variables independientes que discriminarán entre las categorías de la variable dependiente de una manera perfecta; esto le permite al investigador examinar si existen diferencias significativas entre los grupos, en términos de las variables predictoras, también evalúa la precisión de la clasificación[6].

El análisis discriminante crea un modelo predictivo para la pertenencia al grupo. El modelo está compuesto por una función discriminante (o, para más de dos grupos, un conjunto de funciones discriminantes) basada en combinaciones lineales de las variables predictoras que

3 http://www.spssfree.com/curso-de-spss/analisis-descriptivo/escalas-de-medida.html
4 De la fuente F. S. (2011) Análisis discriminante (Universidad autónoma de Madrid)
5 Fuente: (CSIC) Concejo superior de investigaciones científicas.
6 Statistics solutions http://www.statisticssolutions.com/discriminant-analysis/(soluciones estadísticas).

proporcionan la mejor discriminación posible entre los grupos. Las funciones se generan a partir de una muestra de casos para los que se conoce el grupo de pertenencia; posteriormente, las funciones pueden ser aplicadas a nuevos casos que dispongan de mediciones para las variables predictoras pero de los que se desconozca el grupo de pertenencia. La variable de agrupación puede tener más de dos valores. Los códigos de la variable de agrupación han de ser números enteros y es necesario especificar sus valores máximo y mínimo. Los casos con valores fuera de estos límites se excluyen del análisis.

El análisis discriminante se describe por el número de categorías que posee la variable dependiente, de tal manera que cuando esta tiene dos categorías, entonces el tipo utilizado es el análisis discriminante de dos grupos, o si la variable dependiente tiene tres o más de tres categorías, entonces el tipo utilizado es análisis discriminante múltiple. La principal distinción para los tipos de análisis discriminante es que para un grupo de dos, es posible derivar una sola función discriminante. Por otro lado, en el caso del análisis discriminante múltiple, se puede calcular más de una función discriminante[7].

Hay muchas situaciones de nuestra realidad en las que pueden ser importante el análisis discriminante, como por ejemplo, se puede usar para saber si los usuarios de refrescos pesados, medios y ligeros son diferentes en cuanto a su consumo de alimentos congelados; en el campo de la psicología, se puede utilizar para diferenciar entre los compradores de comestibles sensibles al precio y no sensibles al precio en términos de sus atributos o características psicológicas; en el campo de los negocios, se puede utilizar para comprender las características o los atributos de un cliente que posee lealtad de tienda y un cliente que no tiene lealtad de tienda[8], estas, y muchas situaciones más hacen del análisis discriminante una gran ayuda para la solución de problemas donde se requiere clasificar o dar las opciones más acertadas.

1.2.1. Para la clasificación en dos grupos

Veamos la aplicación del Análisis Discriminante a la clasificación de individuos en el caso en que se puedan asignar solamente a dos grupos a partir de p variables discriminadoras. El problema que se había planteado y quedo resuelto por Ronald A. Fisher mediante su función

7 De la fuente F. S. (2011) Análisis discriminante (Universidad autónoma de Madrid)
8 Aldas M. J. (2005) El análisis discriminante. Universidad de Valencia.

discriminante: $D = u_1X_1 + u_2X_2 + u_3X_3 + \ldots + u_pX_p$ por lo tanto las puntuaciones discriminantes son los resultados que se obtienen al dar valores a $X_1, X_2, \ldots, X_p$ en la ecuación anterior. Se trata de obtener los coeficientes de ponderación u_j. Si se considera N observaciones de la función discriminante $D_i = u_1X_{1i} + u_2X_{2i} + u_3X_{3i} + \ldots + u_pX_{pi}$, $\forall i = 1, \ldots, \mathrm{N}$ donde D_i es la puntuación discriminante correspondiente a la observación i−ésima. Función discriminante en forma matricial:

$$\begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{p1} \\ X_{12} & X_{22} & X_{p2} \\ \vdots & \vdots & \vdots \\ X_{1N} & X_{2N} & X_{pN} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}.$$

Según Aldas M. Joaquín (2005), para esto sirve el análisis discriminante. Dada una población, que tenemos dividida en grupos, el análisis discriminante encuentra una función que permite, con un determinado grado de acierto, explicar esa división en grupos (visión explicativa); una vez obtenida, puede utilizarse para clasificar a nuevos individuos en alguno de los grupos en que está dividida la población (visión predictiva).

En los estadísticos empleados podemos encontrar que:

- **Para cada variable**: Medias, desviaciones estándar, ANOVA univariados.
- **Para cada análisis**: M de Box, matriz de correlaciones intra-grupos, matriz de covarianzas intra-grupos, matriz de covarianzas de los grupos separados, matriz de covarianzas total.
- **Para cada función discriminante canónica**: Autovalores, porcentaje de varianza, correlación canónica, Lambda de Wilks, Chi-cuadrado.
- **Para cada paso**: probabilidades previas, coeficientes de la función de Fisher, coeficientes de función no tipificados, Lambda de Wilks para cada función canónica[9].

El análisis discriminante clasifica las observaciones de la muestra en grupos, a partir de la información suministrada por un conjunto de variables (Figura 1).

9 De la fuente F. S. (2011) Análisis discriminante (Universidad autónoma de Madrid)

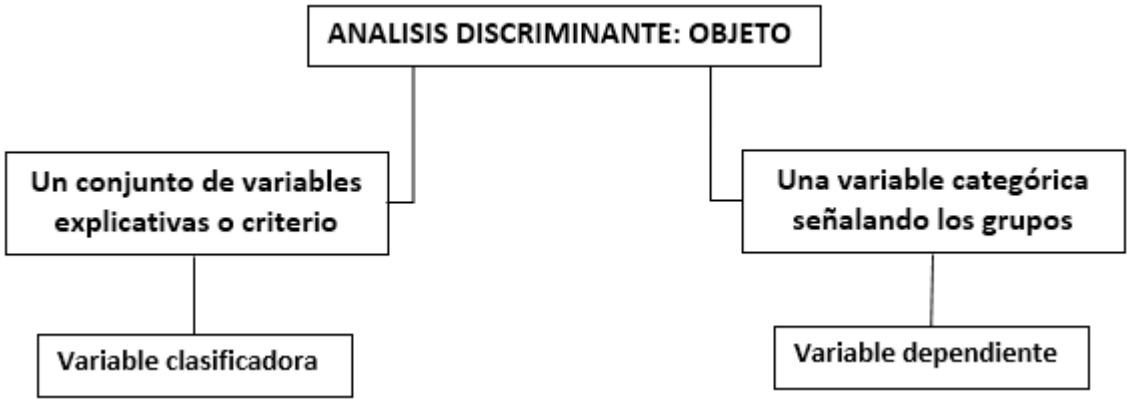


Figura 1: Objeto del análisis discriminante [5]. Editado por el autor.

Ofreciendo una intuición geométrica del análisis discriminante que nos servirá, además, para introducir algunos conceptos necesarios. Supongamos que tenemos una población que puede dividirse en dos grupos, supongamos, también, que queremos ser capaces de explicar esa clasificación atendiendo a una única variable, de esta información podría obtenerse fácilmente lo siguiente (Figura 2).

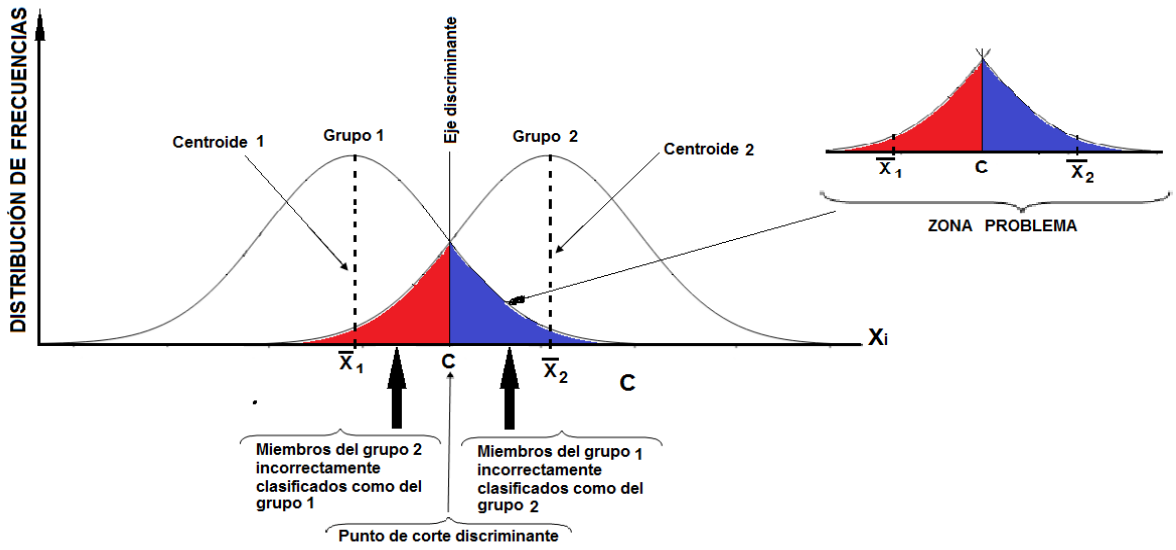


Figura 2: Análisis discriminante y sus funciones de distribución hipotética para dos rupos [4]. Editado y mejorado por el autor.

1.2.2. Los criterios de clasificación en el Análisis discriminante

Hipótesis: Las distribuciones sólo se diferencian por su localización (igual forma y varianza). Se trata de minimizar los errores de clasificación. Si $x_i < C$ se clasifica en el grupo I o Si $x_i > C$ se clasifica en el grupo II. El punto C se denomina punto de corte discriminante:

$$C = \frac{\bar{x}_1 + \bar{x}_2}{2} \quad .$$

La media de ambas medias (C) sería un buen punto de corte como se ilustra en la figura 2. Este criterio, como también se observa en la figura 2, no es infalible, dado que hay elementos del grupo 1 que pueden actuar como del grupo 2 y, por el contrario, elementos del grupo 2 que pueden actuar como del grupo 1. La misión del análisis discriminante es obtener un criterio de clasificación que reduzca ese error; es decir, encontrar una función discriminante que separe lo mejor posible las dos poblaciones.

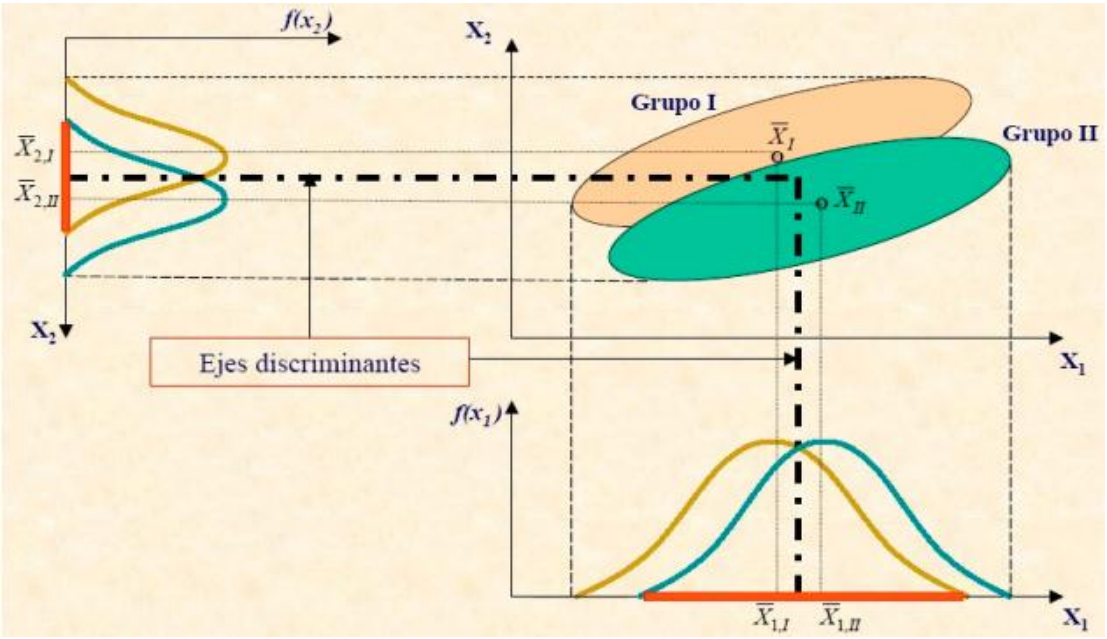


Figura 3. Ejemplo gráfico discriminante, de cómo X_2 discrimina mejor que X_1. [5]

1.2.3. El caso de más de dos grupos

La técnica del análisis discriminante puede utilizarse para efectuar clasificaciones en más de dos grupos. No obstante, cuando se dispone de más de dos grupos de clasificación, la interpretación de los resultados cambia ligeramente.

Con más de dos grupos es posible obtener más de una función discriminante. En concreto, es posible obtener tantas como número de grupos menos uno (a no ser que el número de variables independientes sea menor que el número de grupos, en cuyo caso el número de posibles funciones discriminantes será igual al número de variables menos uno).

Las funciones discriminantes se extraen de manera jerárquica, de tal forma que la primera función explica el máximo posible de las diferencias entre los grupos, la segunda función explica el máximo de las diferencias todavía no explicadas, y así sucesivamente hasta alcanzar el 100% de las diferencias existentes. Esto se consigue haciendo que la primera función obtenga el mayor cociente entre las sumas de cuadrados inter-grupos e intra-grupos. La segunda, el siguiente mayor cociente entre ambas sumas de cuadrados, etc.

Además, las funciones resultantes son ortogonales o independientes entre sí. En el caso de tres grupos, por ejemplo, el efecto final de esta independencia es que la primera función intenta discriminar lo mejor posible entre dos de los grupos y, la segunda, entre los dos grupos que aún se encuentren más próximos.

Las etapas del análisis discriminante son[10]:

1. Planteamiento del problema (se formula antes de realizar).
2. Selección de variables dependientes e independientes.
3. Selección del tamaño muestral.
4. Comprobación de las hipótesis de partida.
5. Estimación del modelo.
6. Los coeficientes de función discriminante son estimados.
7. Validación de las funciones discriminantes (es la determinación del significado de estas funciones discriminantes).
8. Contribución de las variables a la capacidad discriminante.

[10] Análisis discriminante. Universidad de Córdoba (mejorado por el autor).

9. Valoración de la capacidad predictiva.
10. Selección de variables.
11. Se debe interpretar los resultados obtenidos.
12. El último y más importante paso es evaluar la validez.

1.3. Métodos kernel[11]

Son métodos que se utilizan en un conjunto de datos que provienen de una distribución continua y desconocida para aproximar a una función; en el aprendizaje automático, los métodos kernel son una clase de algoritmos para el análisis de patrones, cuyo miembro más conocido es la máquina de vectores de soporte (SVM). La tarea general del análisis de patrones es encontrar y estudiar tipos generales de relaciones (por ejemplo: clústeres, clasificaciones, componentes, principales, correlaciones) en conjuntos de datos. Para muchos algoritmos que resuelven estas tareas, los datos en su representación normal deben transformarse explícitamente en representaciones de vectores de características a través de un mapa de características especificado por el usuario, en contraste, los métodos kernel requieren solo un kernel especificado por el usuario, es decir, una función de similitud sobre pares de puntos de datos en representación real.

Los métodos kernel deben su nombre al uso de funciones kernel, que les permiten operar en un espacio de características implícitas de alta dimensión sin tener que calcular las coordenadas de los datos en ese espacio, sino simplemente calculando los productos internos entre las imágenes de todos los pares de datos en el espacio de características. Los kernel más comunes son:

- Lineal: $K(x_i, x_j) = x_i^T x_j$
- Gaussiano: $K(x_i, x_j) = \exp\left(\frac{-\|xi - xj\|^2}{2\sigma^2}\right)$
- Polinómica: $K(x_i, x_j) = (x_i^T x_j + 1)^n$

Otras funciones kernel son:

- Neuronal (Sigmoid, Tanh) Kernel: $K(x_i, x_j) = \tanh(ax_i * x_j + b)$
- Anova Kernel: $K(x_i, x_j) = (\sum_i^n exp\,(-\gamma\,(x_i - x_j))^d$

[11] Manel M. R. (2008) Introducción a los métodos Kernel, Universidad Autónoma de Madrid.

Fourier series Kernel: $K(x_i, x_j) = \frac{sen(N+\frac{1}{2})(x_i - x_j)}{sen(\frac{1}{2}(x_i - x_j))}$

Spline Kernel: $K(x_i, x_j) = \sum_{r=0}^{a} x_i^r x_j^r \sum_{s=1}^{N} (x_i - t_s)^a + (x_j - t_s)^a$

Esta operación suele ser computacionalmente más económica que el cálculo explícito de las coordenadas. Este enfoque se llama el " truco del kernel". Las funciones del kernel se han introducido para datos de secuencia, gráficos , texto, imágenes, así como vectores.

Los algoritmos que pueden operar con kernels, máquinas de vectores de soporte (SVM), procesos Gaussianos , análisis de componentes principales (PCA), análisis de correlación canónica, regresión de crestas, agrupamiento espectral, filtros adaptativos lineales y muchos otros. Cualquier modelo lineal puede convertirse en un modelo no lineal aplicando el truco del kernel al modelo: reemplazando sus características (predictores) por una función.

La mayoría de los algoritmos del núcleo se basan en la optimización convexa o en problemas propios y están estadísticamente bien fundamentados. Típicamente, sus propiedades estadísticas se analizan usando la teoría del aprendizaje estadístico.

1.3.1. Motivación y explicación informal del kernel

Los métodos kernel pueden considerarse aprendices basados en instancias: en lugar de aprender un conjunto fijo de parámetros correspondientes a las características de sus entradas, en su lugar "recuerdan" la i- ésima muestra de entrenamiento (x_i, y_i) y aprende para ello un peso correspondiente w_i. La predicción para entradas no etiquetadas, es decir, aquellas que no están en el conjunto de entrenamiento, se trata mediante la aplicación de una función de similitud k, llamado kernel, entre la entrada no etiquetada x' y cada una de las entradas de entrenamiento x_i. Por ejemplo, un clasificador binario del kernel generalmente calcula una suma ponderada de similitudes.

$$\hat{y} = sgn \sum_{i=1}^{n} w_i \, y_i K(x_i, x'),$$

$\hat{y} \in \{-1, +1\}$ es la etiqueta predicha del clasificador binario del kernel para la entrada no etiquetada x' cuya verdadera etiqueta oculta es de interés; $K: x * x \longrightarrow \mathbb{R}$ es la función del núcleo que mide la similitud entre cualquier par de entradas $(x, x') \in X$.

1. La suma varía sobre los n ejemplos etiquetados $\{(x_i, y_i)\}_{i=1}^{n}$ en el conjunto de entrenamiento del clasificador, con $y_i \in \{-1, +1\}$.
2. El $\mathcal{W}_i \in \mathbb{R}$ son los pesos para los ejemplos de entrenamiento, según lo determinado por el algoritmo de aprendizaje; la función de signo sgn determina si la clasificación predicha $\hat{y}$ sale positivo o negativo.

1.3.2. Algunos aspectos históricos de los métodos kernel.

Los clasificadores de kernel se describieron en la década de 1960, con la invención del kernel perceptron. Se convirtieron en una gran prominencia con la popularidad de la máquina de vectores de soporte (SVM) en la década de 1990.

En la clasificación estadística encontramos el kernel de Fisher, llamado así por Ronald Fisher, es una función que mide la similitud de dos objetos sobre la base de conjuntos de mediciones para cada objeto y un modelo estadístico. En un procedimiento de clasificación, la clase para un nuevo objeto (cuya clase real es desconocida) puede estimarse minimizando, a través de las clases, un promedio de la distancia del kernel de Fisher desde el nuevo objeto a cada miembro conocido de la clase dada.

El kernel de Fisher se introdujo en 1998. Combina las ventajas de los modelos estadísticos generativos (como el modelo oculto de Markov) y los de los métodos discriminativos (como las máquinas de vectores de soporte):

1. Los modelos generativos pueden procesar datos de longitud variable (agregar o eliminar datos está bien respaldado).
2. Los métodos discriminativos pueden tener criterios flexibles y producir mejores resultados.

1.3.3. Kernel, definición y ejemplo[12]

Las representaciones Kernel ofrecen una solución alternativa al proyectar los datos en un espacio de características de alta dimensión para aumentar la potencia de cálculo de las máquinas de aprendizaje lineal. La ventaja de utilizar las máquinas en la representación dual deriva del hecho de que, en esta representación, el número de parámetros ajustables no depende del número de atributos que se utilizan. Al reemplazar el producto interno con una función de kernel elegida apropiadamente, se puede realizar implícitamente una función no lineal a un espacio de características de alta dimensión sin aumentar el número de parámetros ajustables, siempre que el núcleo calcule el producto interno de los vectores de características correspondientes a las dos entradas. Por tanto, el problema consiste en elegir un kernel adecuado para la SVM. El principal interés de los kernel, en el contexto de SVM, es que lo que se ha visto en el caso de separación lineal también se aplica fácilmente a la separación no lineal mediante su uso[13].

Para construir un estimador no lineal, debemos transformar los datos de entrada no linealmente. La transformación no lineal implica una correspondencia hacia un espacio de mayor dimensión, posiblemente infinita:

$$\varphi: \mathbb{R}^n \longrightarrow \mathcal{H}$$
$$x \longrightarrow \varphi(x).$$

Un ejemplo de transformación no lineal a un espacio de mayor dimensionalidad es una transformación polinómica: sea un conjunto de datos unidimensional x_i. Aplicamos la siguiente transformación no lineal: $\varphi(x) = \{x^2, \sqrt{2}x, 1\}^T \in \mathbb{R}^3$.

La pregunta es: existe un producto escalar en ese espacio que pueda ser expresado como función de los datos de entrada x . El producto escalar explícito es en el método Kernel es:

$$\varphi(x_1)^T \varphi(x_2) = \{x_1^2, \sqrt{2}x_1, 1\}\{x_2^2, \sqrt{2}x_2, 1\}^T = x_1^2 x_2^2 + 2x_1 x_2 + 1 .$$

Que puede ser escrito como

$$\varphi(x_1)^T \varphi(x_2) = x_1^2 x_2^2 + 2x_1 x_2 + 1 = (x_1 x_2 + 1)^2.$$

• Existe una expresión del producto escalar en función del espacio de entrada.

[12] Martínez R. Manuel (2008). Introducción a los métodos Kernel. Universidad Autónoma de Madrid.
[13] López D. Ana (2017). Fundamentos Matemáticos de los Métodos Kernel para Aprendizaje Supervisado. Facultad de matemáticas. Universidad de Sevilla. España. Pag. 40

• Ese producto escalar se denomina Kernel, y el espacio de mayo r dimensionalidad (espacio de características) es un espacio de Hilbert (Reproducing Kernel Hilbert Space, RKHS).

• No necesitamos la expresión de las componentes del vector en el espacio de características. Esta transformación incrementa la posibilidad de que haya separabilidad lineal. Este es un ejemplo de correspondencia en un espacio de Hilbert. En una dimensión, no es posible clasificar los datos linealmente.

1.3.4. El truco de los Kernels.

El hecho de necesitar sólo los productos escalares nos permite reproducir cualquier algoritmo lineal en un espacio de Hilbert; esto es, existe una versión no lineal de cualquier algoritmo lineal basado en datos. Si encontramos una transformación no lineal $\varphi(x)$ a un espacio de mayor dimensionalidad provisto de un producto escalar que puede ser expresado como (Kernel):

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) .$$

Entonces podremos construir una versión no lineal del mismo algoritmo donde la transformación no lineal es φ. Este es el truco de los Kernels.

En el " truco del kernel" las funciones del kernel se introducen para datos de secuencia, gráficos , texto, imágenes, etc. Los algoritmos que pueden operar con kernels, procesos gaussianos, lineales o polinómicas, de correlación canónica o agrupamiento espectral , filtros adaptativos lineales y muchos otros. Todo esto mediante la premisa que cualquier modelo lineal puede convertirse en un modelo no lineal aplicando el truco del kernel al modelo reemplazando sus características (predictores) por una función.

Recordemos que no se necesitan los vectores, así que no debemos preocuparnos por la dimensión del espacio en cuanto a costo computacional. Lo que necesitamos es conocer el kernel.

- **Aplicaciones kernel:** Los kernel mediante el uso de sus fórmulas pueden servir en estadística para la organización y clasificación de datos, además, dentro de este proceso podemos encontrar aplicaciones de los kernel los cuales están dados en: Recuperación de información, clasificación y recuperación de imágenes y transformaciones no lineales.

1.3.5. Criterio de clasificación con el kernel

Para la clasificación de individuos con la función kernel se utiliza la media de los grupos y se relaciona con cada dato especifico y a esto se le estima la varianza en modelación (programa Excel) para verificar cual es la varianza más efectiva de acuerdo al porcentaje de los individuos clasificados. El valor final obtenido se encuentra en un intervalo entre 0 y 1, en este caso se asume la similaridad como la cercanía al grupo o específicamente al centroide; o sea que, mientras más cercano este el valor obtenido, de uno (1) demuestra la cercanía a ese grupo.

Para clasificar las variables se utiliza la desviación estándar de cada variable en cada grupo y se relaciona con el total ponderado, a esto se le estima la varianza en modelación (programa Excel). El valor final obtenido se encuentra en un intervalo entre 0 y 1, donde cero (0) indica que no hay similaridad o la similaridad es nula, y el uno (1) indica similaridad perfecta, esto se asemeja al estadístico Lambda de Wilks para identificar el poder discriminante de la variable. Esto permite seleccionar las variables más importantes, puesto que los valores más separados de uno (1) indican más discriminación.

Para obtener los valores anteriormente referenciados se hace uso de una función kernel, para este caso se usa la función kernel Gaussiana:

$$K\left(x_i, \bar{x}_j\right) = \exp\left(\frac{-\left\|x_i - \bar{x}_j\right\|^2}{2\sigma^2}\right).$$

Donde, para efectos del estudio x_i representa el dato del individuo o de la variable en el grupo, además, dependiendo el ensayo puede ser la desviación estándar de la variable en el grupo; $\bar{x}_j$ representa la media del grupo, la media de la variable o la media de las desviaciones estándar en cada variable; σ^2 representa la varianza estimada o de ensayo; K representa el valor del kernel en cada situación.

Dentro del campo del aprendizaje supervisado, encontramos dos tipos de problemas: los aprendizaje multitarea mediante procesos gaussianos para clasificación en problemas de regresión, para los que el objetivo es aprender una función continua, y los problemas de

clasificación, en los que se pretende categorizar observaciones en una clase o grupo dentro de dos o más posibles clases.

Por último, cabe mencionar una propiedad importante de los métodos kernel. Dentro del aprendizaje supervisado, encontramos fundamentalmente dos tipos de métodos: paramétricos y no paramétricos. Los primeros asumen a priori una distribución para los datos, que es función de unos parámetros que se determinan a partir de los datos. Los métodos kernel, son métodos no paramétricos ya que no asumen que la distribución a la que se ajustan los datos tienen un número fijo de parámetros. Sin embargo, es necesario elegir un kernel o función de covarianza que determinará la solución al problema. En este sentido, estudiaremos un kernel con la propiedad de ser muy expresivo, es decir, que pueden describir patrones muy diferentes en los datos ya que puede aproximar arbitrariamente kernel muy diferentes[14].

Las funciones kernel representan una suposición básica en los métodos kernel, incluyendo los procesos gaussianos: dados dos datos x_1 y x_2 esperamos que la función f que aprenda nuestro algoritmo tome valores similares para dichos datos, es decir que $f_1 = f(x_1) \approx f(x_2) = f_2$.

La función kernel más sencilla que podemos construir se hace a partir de la identidad $\varphi(x) = x$. En este caso el kernel es $k(x, x') = x^T x'$ y por tanto el producto escalar canónico es una función kernel válida.

Una idea interesante es el truco conocido habitualmente como kernel trick: si tenemos un algoritmo en el que los datos de entrada lo hacen en forma de producto escalar, podemos reemplazar el producto escalar por un kernel cualquiera.

A partir de la definición que hemos dado de función kernel, podemos deducir una propiedad muy sencilla pero muy importante: cualquier función kernel k es simétrica (es decir, $k(x, x') = k(x', x)$). Esta propiedad nos puede ayudar a ganar intuición sobre lo que es una función kernel $k(x, x')$ es una medida de la similitud entre los vectores x y x'.

[14] R. P. Wilson, A. G. Adams. Gaussian process kernels for pattern discovery and extrapolation. ICML, Pag. 1067 - 1075, 2013.

1.4. Algunos teóricos del rendimiento académico

Martínez Hernández y Núñez (1987), expresa que: "El bajo rendimiento académico se debe a múltiples factores como falta de motivación y comunicación de los padres de familia hacia el estudiante, considerado que estos problemas son similares en todas partes; además, es necesario crear estrategias.

Jiménez (2000), delimita el rendimiento escolar como un nivel de conocimientos que se muestran en un área o asignatura contrastado con la norma de edad y nivel académico. Montero &Villalobos (2007), definen el rendimiento académico desde un conjunto de factores que afectan el resultado académico en donde intervienen aspectos de orden sociodemográfico, psicosociales, pedagógicos, institucionales y socioeconómicos.

No obstante, el éxito académico está determinado por múltiples factores que van desde habilidades cognitivas, intereses, motivación, autoconcepto, ansiedad, hábitos de estudio, contexto sociohistórico, dinámica familiar, salud, ambiente escolar, influencia de padres y compañeros, escolaridad de los padres, hasta variables relacionadas con los programas, el currículo, las características de quien enseña y cómo lo hace; además de una gran cantidad de factores externos (Córtes, 2008).

Teniendo en cuenta las definiciones de los autores anteriores, el rendimiento académico se puede ver como el nivel de conocimientos adquiridos por un estudiante, los cuales están influenciados por diferentes factores tanto académicos como sociales, económicos, psicológicos, entre otros[15]. Pero dentro de todos estos factores, la participación permanente y activa de los padres de familia es fundamental ya que los estudiantes en esta época muestran demasiada tranquilidad y poca responsabilidad la cual puede disminuir con una buena acción responsable de los padres.

[15] Marquín T. María J. (2017). Tesis Predicción del rendimiento académico mediante técnicas del análisis multivariado en la asignatura de algebra lineal. Facultad de ingeniería industrial. Universidad Tecnológica de Pereira. Pereira, Risaralda. Pag 15-16.

2. METODOLOGÍA

2.1. Característica de la población objeto de estudio

El establecimiento educativo donde se lleva a cabo la aplicación del trabajo de tesis es la Institución Educativa Técnica Occidente de la ciudad de Tuluá (Valle del Cauca) cuya modalidad es técnico empresarial y cuenta con una sede principal. Durante el año lectivo 1996-1997, se adopta la modalidad de gestión empresarial enfocada desde el horizonte de autogestión, buscando que los estudiantes piensen en la posibilidad de proyectar su propio negocio.

Atendiendo una población de estratos 1, 2 y 3, con niños cuyos padres se dedican al corte de caña generalmente, madres empleadas de servicio doméstico, economía informal entre otros, donde su nivel educativo en un alto porcentaje es bajo y medio. En la actualidad la Institución Educativa cuenta con un cuerpo docente integrado por 111 profesores y alrededor de 3200 estudiantes aproximadamente de preescolar a grado 11 en jornada mañana y tarde.

La población objeto de estudio a la cual se le implementa el trabajo de tesis, son los estudiantes correspondientes a los grados 8 (octavos) de educación básica para 5 cursos de 8-1 a 8-5, abarcando 184 estudiantes, actualmente en la institución educativa se tienen nueve grupos en este nivel, cada grupo aproximadamente de 35 a 40 estudiantes, con edades que oscilan entre los 13 y 16 años y que habitan en la zona noroccidental del municipio de Tuluá y en un bajo porcentaje estudiantes de otras zonas de la ciudad.

2.2. Variables aplicadas en el proceso

Para este trabajo se tuvieron en cuenta quince (15) variables que se agruparon de la siguiente forma:

2.2.1. Rendimiento.

1. Variable Grupo: **Nivel** (Nota promedio: Nivel de competencia).

Nota: Se toma la calificación mínima desde 1,0 por políticas de la institución, en donde se dice que la calificación mínima para un estudiante debe ser 1,0.

2. Pérdidas (Asignaturas perdidas): Nota: **Bueno** es el estudiante que no perdió asignaturas, **Regular** es el estudiante que perdió 1 o 2 asignaturas, y según políticas de la institución puede aprobar el grado pero debe realizar actividades de apoyo al iniciar el siguiente año lectivo, **Deficiente** es el estudiante que perdió 3 o más asignaturas, y según políticas de la institución no pude aprobar el grado, pero si perdió 3 o 4 asignaturas puede acceder a realizar una actividades de apoyo especial para una promoción anticipada al iniciar el siguiente año lectivo.

Tabla 1. Variables de rendimiento (independiente asignaturas perdidas y dependiente nota promedio)

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
RENDIMIENTO	1. Nota promedio: Nivel de competencia	Cualidad producto de la calificación promedio obtenida en las notas de las áreas de todo el pensum académico en grado octavo. Se expresa en la variable con una escala de 1 a 10 puntos.	**Nivel** Variable Grupo	1. Bajo 2.Básico 3. Alto 4. Superior	1. (1,0 - 5,9) 2. (6,0 - 7,9) 3. (8,0 - 8,9) 4. (9,0 - 10,0)
	2. Asignaturas perdidas	Es una cualidad del estudiante que se obtiene producto de la cantidad de asignaturas perdidas o no haber perdido asignaturas.	**Pérdidas**	1. Deficiente 2.Regular 3. Bueno	1. (3 o más asignaturas perdidas) 2. (1 ó 2 asignaturas perdidas) 3. (No tiene asignaturas perdidas)

2.2.2. Responsabilidad.

3. **Tiempodedif (**Dedicas tiempo a estudiar por fuera de clase)
4. **Entregatareas** (Entrega a tiempo sus tareas). Cualidad del estudiante que refleja la responsabilidad para entregar tareas a tiempo.

Tabla 2. Variables de responsabilidad, independientes.

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
RESPONSABILIDAD	3. Dedicas tiempo a estudiar por fuera de clase)	Cualidad del estudiante que refleja la dedicación por fuera de la clase	**Tiempodedif**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría
	4. Entrega a tiempo sus tareas.	Cualidad del estudiante que refleja la responsabilidad para entregar tareas a tiempo	**Entregatareas**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.2.3. Acompañamiento.

5. **Padrespend** (Sus padres están al pendiente del desarrollo de sus actividades académicas)
6. **Vivepadres** (Vives con tus padres).

Tabla 3. Variables de acompañamiento, independientes.

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
ACOMPAÑAMIENTO	5. Sus padres están al pendiente del desarrollo de sus actividades académicas.	En esta variable se identifica que tanto los padres están pendientes de las actividades académicas sus hijos y si apoyan en buena forma el proceso académico	**Padrespend**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría
	6. Vives con tus padres.	Aquí se pretende saber si el estudiante vive con los padres o con otro tipo de acudiente, aspecto vital en el rendimiento académico.	**Vivepadres**	1. Un conocido. 2. Un familiar. 3. Uno de los padres. 4. Con los dos padres.	Los mismos valores de 1 a 4 de acuerdo a la elección de la categoría

2.2.4. Motivación.

7. **Ambientefam** (Hay buen ambiente en su casa para desarrollar sus actividades académicas).
8. **Educacionrec** (Crees que la educación que estas recibiendo en el colegio es buena).

Tabla 4. Variables de motivación, independientes.

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
MOTIVACIÓN	7. Hay buen ambiente en su casa para desarrollar sus actividades académicas.	Esta variable muestra el nivel de convivencia que pueden estar pasando los estudiantes en sus casas	**Ambientefam**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría.
	8. Crees que la educación que estas recibiendo en el colegio es buena.	Aceptación de los procesos educativos que recibe el estudiante	**Educacionrec**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.2.5. Expectativa

9. **Tegustacol (**Te gusta el colegio).

10. **Haspenfu** (Has pensando en tu futuro).

Tabla 5. Variables de expectativa, independientes.

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
EXPECTATIVA	9. Te gusta el colegio.	Esta variable muestra el gusto del estudiante por los elementos, normas y estructura de la institución educativa.	**Tegustacol**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría
	10. Has pensando en tu futuro.	Aceptación, deseos y aspiraciones del estudiante a mejorar su calidad de vida.	**Haspenfu**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.2.6. Convivencia

11. Vicbullyng (Es víctima de bullyng en el colegio)

Tabla 6. Variable de convivencia, independiente.

CONVIVENCIA	11. Es víctima de bullyng en el colegio.	En esta variable se identifica el trato a los estudiantes por parte de sus compañeros	**Vicbullyng**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.2.7. Nivel familiar

12. Nivelespa (Cuál es el nivel de escolaridad de tus padres o tu acudiente).

13. Cartracasa (Le ponen excesiva carga de trabajo en su casa).

Tabla 7. Variables de nivel familiar, independientes.

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN
NIVEL FAMILIAR	12. Cuál es el nivel de escolaridad de tus padres o tu acudiente.	Esta variable mide el nivel de estudio de los padres o acudientes en el estudiante evaluado.	**Nivelespa**	1. No tiene estudio. 2. Primaria. 3. Bachillerato. 4. Pregrado. 5. Posgrado.	Los valores de 1 a 5 de acuerdo a la elección de la categoría presenciada en el estudiante.
	13. Le ponen excesiva carga de trabajo en su casa.	Esta variable permite determinar si el estudiante en su casa es sometido a largas horas de trabajo que no permitan su avance académico.	**Cartracasa**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.2.8. Cumplimiento de normas

14. Respetocodoc (Tiene respeto por sus compañeros y docentes).

15. Pactoconv (Reconoce el pacto de convivencia y lo cumple).

Tabla 8. Variables de cumplimiento de normas, independientes.

CUMPLIMIENTO DE NORMAS	14. Tiene respeto por sus compañeros y docentes.	Variable que identifica el respeto del estudiante hacia la comunidad educativa.	**Respetocodoc**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría
	15. Reconoce el pacto de convivencia y lo cumple.	Variable que identifica si el estudiante reconoce los acuerdos estudiantiles dados en su matrícula.	**Pactoconv**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría

2.3. Aspectos importantes para interpretación del análisis estadístico

2.3.1. Interpretación del análisis discriminante (AD)

En el análisis discriminante AD, es una técnica multivariante orientada fundamentalmente a lograr objetivos básicos como[16]:

Explorar y analizar las posibles diferencias que puedan existir entre g poblaciones excluyentes, previamente definidas por el investigador, en base a las diferencias que puedan presentar en las p variables medidas. Se trata de hallar funciones que dependan de esas p variables originales que separen los g grupos tanto como sea posible.

A partir del criterio de discriminación obtenido se puede proceder a incluir un nuevo elemento en algunos de los grupos formados. Este es el caso de los individuos que no se les conoce a priori el grupo al cual pertenece, entonces el AD permite clasificarlos sobre la base de ecuaciones matemáticas, derivadas del análisis de los casos con pertenencia conocida. Para el estudio se nombran como variables discriminantes aquellas variables utilizadas en el análisis estadístico. Estas deben ser medidas en la escala de intervalo o razón para que las medias y varianzas puedan ser calculadas e interpretadas.

Una variable que es considerada como original no puede ser combinación lineal de otras variables discriminantes ya que se tendría una redundancia en la información. Una combinación lineal es la suma de una o más variables que pueden haber sido ponderadas por términos constantes. Del mismo modo, dos variables que están perfectamente correlacionadas no pueden ser usadas al mismo tiempo.

2.3.2. Supuestos del análisis discriminante.

En este estudio se sigue el esquema tradicional del AD que exige los siguientes supuestos:

Cada grupo debe ser considerado como una muestra extraída de una población normal multivariante.

16. Garnica O. Elsy, González M. Pilar, Díaz de Pascual Amelia y Enrique Torres L. (1991) Análisis discriminante: Estudio del rendimiento estudiantil. Universidad de Los Andes, Instituto de Investigaciones Económicas y Sociales

Las matrices de covarianzas poblacionales de todos los grupos deben ser iguales (esto puede probarse con el test para la homogeneidad de varianzas de Box). Este supuesto permite la simplificación de la fórmula lineal discriminante y la adecuada interpretación de los resultados de los test de significación. A veces estos supuestos son difíciles de hallar. Algunos autores, entre ellos. Lachenbruch han demostrado que el AD es una técnica robusta que puede tolerar ciertas desviaciones de estos supuestos (Lachenbruch, P. A., 1975).

Si se cumple el primer supuesto (normalidad), las funciones lineales discriminantes minimizan la probabilidad de mala clasificación.

Existen diversas reglas de clasificación, entre ellas: la regla de clasificación de máxima verosimilitud, la regla de clasificación de Bayes, regla de clasificación generalizada de Bayes y la función discriminante cuadrática en poblaciones normales; además, se disponen de diversos métodos de evaluación de las funciones de clasificación: holdout, resustitución técnica, jackknife, entre otros, (Márquez, 1989).

Cuando el supuesto de normalidad es violado, el cálculo de probabilidades no es exacto. En este caso, al tener probabilidades aproximadas, los resultados deben interpretarse con cuidado ya que puede haber una reducción en eficiencia y seguridad, sobre todo en el caso de muestras muy pequeñas.

En los casos frontera, los pequeños errores debido a la violación del supuesto de normalidad podrían causar una clasificación incorrecta. Por ejemplo, si un caso tiene una probabilidad de 0,95 de pertenecer al grupo 1 y 0,05 de pertenecer al grupo 2, no importa la imprecisión debida a la violación de los supuestos, ya que la decisión para asignar el caso al grupo 1 será posiblemente correcta; si el caso tiene una probabilidad de 0,51 de pertenecía del grupo 1 y de 0,49 para el grupo 2, se debe ser cauteloso al tomar una decisión.

Es frecuente conseguir que el requerimiento de igualdad de matrices de covarianzas de los grupos, no se cumpla. Cuando este supuesto es violado se presentan distorsiones en la función discriminante canónica y la ecuación de clasificación. Cuando se determina la existencia de diferencias significativas, aún es posible utilizar, con buenos resultados la función lineal discriminante si las matrices de covarianzas no son muy diferentes. Si se ha comprobado que las matrices de covarianzas de grupos son muy diferentes, se sugiere usar estas matrices para calcular la probabilidad de pertenencia de grupo. Este análisis se denomina discriminación cuadrática.

En el AD se obtiene una combinación de las variables independientes, para cada caso o individuo, denominado score (D_i):

$$D_i = u_1 X_{1i} + u_2 X_{2i} + u_3 X_{3i} + \ldots + u_p X_{pi},$$

que resume la información contenida en todas las variables en un sólo índice. Además, esta combinación lineal se utiliza para asignar nuevos casos o individuos a los grupos ya formados.

Por lo tanto las puntuaciones discriminantes son los resultados que se obtienen al dar valores a $X_1, X_2, \ldots, X_p$ en la ecuación anterior. Se trata de obtener los coeficientes de ponderación u_j. Si se considera N observaciones de la función discriminante.

Si los puntajes o scores de las funciones discriminantes están distribuidos normalmente para cada uno de los grupos observados y los parámetros de la distribución pueden ser estimados, entonces es posible calcular la probabilidad condicional de obtener un score D_i dado que el individuo en cuestión sea miembro del grupo G_i (para $i = 1,.,g$), es decir $P(D_i / G_i)$. También es posible calcular, por otra parte, la probabilidad a priori $P(Gi)$, la cual es un estimador de máximo verosímil de la probabilidad de que un individuo j pertenezca a un grupo particular G_i, cuando no hay información disponible. Con el conocimiento de las probabilidades condicionales y a priori es posible calificar individuos en uno de los grupos, utilizando la Regla de Bayes:

$$P\left(\frac{G_i}{D_i}\right) = \frac{P\left(\frac{D_i}{G_i}\right) P(G_i)}{\sum_{i=1}^{g} P\left(\frac{D_i}{G_i}\right) P(G_i)},$$

donde:

$P\left(\frac{G_i}{D_i}\right)$: Probabilidad de que un individuo pertenezca al grupo i dado que tiene un score D_i.

$P\left(\frac{D_i}{G_i}\right)$: Probabilidad de que un individuo tenga un score D, dado que pertenece al grupo i.

$P(G_i)$: Probabilidad de que un individuo pertenezca al grupo i.

El porcentaje de correcta clasificación es un indicador de gran utilidad cuando la investigación se centra en la búsqueda de una descripción razonable del mundo real. Si este porcentaje es alto, no importa la violación de los supuestos. Pero, si el porcentaje de correcta clasificación es bajo, no se puede asegurar si esto se debe a la violación de los supuestos o al uso

de variables discriminantes débiles. El centroide de un grupo es un punto cuyas coordenadas son los promedios, en el grupo, de cada una de las variables discriminantes.

En la interpretación de la función discriminante canónica es de gran utilidad analizar la posición relativa de los datos respecto a los centroides.

La correlación entre una variable y la función discriminante, denominada coeficiente de estructura total, es el coseno del ángulo formado por el punto variable y la función. Esta correlación se utiliza para determinar el peso que tiene esa variable en la función discriminante.

Cada una de las funciones discriminantes está asociada a un valor propio (λ). Los valores propios no tienen interpretación directa pero la magnitud de éste indica el poder discriminatorio de la función. Ordenando los valores propios en forma descendente se obtienen las funciones discriminantes organizadas según su poder discriminatorio. En otras palabras, la función con mayor valor propio es el discriminador más poderoso, mientras que la función con el menor valor propio es la más débil. Para comparar los valores propios se acostumbra transformarlos en proporciones relacionados con la suma total de ellos (varianza total). Este valor indica si la función es fuerte o débil relacionada con las demás.

El número de funciones discriminantes del análisis es el mínimo entre p (número de variables) y g (número de grupos menos uno), o sea, $min\ (g-1,\ p)$.

- Distribución normal: Se asume que los datos de las variables representan una muestra proveniente de una distribución normal multivariante.
- Homogeneidad de varianzas y covarianzas: las matrices de varianzas-covarianzas intragrupos deben ser iguales en todos los grupos. Para comprobar esto se puede utilizar una prueba de M de Box, esta prueba tiene como hipótesis nula (H_0) que las matrices de covarianzas son iguales, el valor obtenido se aproxima a una F de Snedecor. Si p < 0.05 se rechaza H_0.

2.3.3. Estadísticos

Para probar la significancia estadística de las funciones discriminantes se examinan los residuales de la discriminación "a priori". Si la discriminación residual es pequeña, no tiene sentido utilizar las restantes funciones aún si ellas existen matemáticamente. La Lambda de Wilks (Λ) es una medida de la diferencia entre los grupos, respecto a las variables discriminadas, expresa la proporción de variabilidad total no debida a las diferencias entre los grupos; permite

contrastar la hipótesis nula de que las medias multivariantes de los grupos (los centroides) son iguales. Este estadístico Λ está definido como:

$$\Lambda = \frac{Suma\ de\ los\ cuadrados\ dentro\ de\ los\ grupos}{Suma\ de\ cuadrados\ totales},$$

el cual toma valores entre cero (0) y uno (1) y su interpretación está en relación inversa al estadístico F, el cual está definido como:

$$F = \frac{Cuadrado\ medio\ entre\ los\ grupos}{Cuadrado\ medio\ dentro\ de\ los\ grupos},$$

y también indica el grado de influencia de cada variable explicativa por separado sobre la variable que hay que explicar, además, expresa su poder discriminante.

Los valores de Lambda de Wilks cerca de cero denotan alta discriminación. Esto quiere decir que los centroides están muy separados entre sí. A medida que la Lambda de Wilks se acerca a uno el poder discriminatorio de la función se hace más débil. Cuando tiene exactamente el valor de uno, los centroides de los grupos son idénticos (no hay diferencias entre los grupos).

Como la distribución de la Lambda de Wilks puede ser aproximada a una distribución χ^2 (Chi- cuadrada), el nivel de significación se puede determinar comparando el valor calculado u observado con el valor tabulado de esta distribución.

El AD no es un análisis causal porque las variables no se definen como dependientes o independientes. Si en determinado estudio se define la variable grupo como independiente de las variables discriminantes se tiene una situación análoga al análisis de regresión múltiple.

Para la determinación del nivel explicativo de cada función discriminante y antes de interpretar estas funciones, tenemos que asegurarnos de que su nivel explicativo es elevado, es decir, de que ayudan realmente a diferenciar los grupos de población analizados. Con esta finalidad, se utilizan los parámetros estadísticos siguientes:

a) El valor propio λ (ratio): asociado a cada función discriminante $\lambda = \frac{(SC_{interg})}{(SC_{intrag})}$. No olvidemos que las funciones discriminantes se obtienen maximizando este ratio; así, valores propios elevados implican funciones discriminantes con un poder explicativo elevado.

b) El porcentaje de varianza entre grupos: explicada por cada función discriminante, se calcula en función del valor propio asociado a cada función discriminante. Si λ_r es el valor propio asociado a la función discriminante D_r, el porcentaje de varianza entre grupos explicada por D_r, donde **R** es el número total de funciones discriminantes, es el siguiente:

$$V(D_r) = \left(\frac{\lambda_r}{\sum_{r=1}^{t} \lambda_r}\right) * 100.$$

✓ **La correlación canónica:** Como medida de evaluación de la bondad de ajuste se utiliza el coeficiente eta cuadrado η^2, que es el coeficiente de determinación obtenido al realizar la regresión entre la variable dicotómica, que indica la pertenencia al grupo, y las puntuaciones discriminantes. A la raíz cuadrado de este coeficiente se le denomina correlación canónica.

$$\overbrace{\eta^2 = \frac{\lambda}{1+\lambda}}^{Eta\ cuadrado} \qquad \overbrace{\eta = \sqrt{\frac{\lambda}{1+\lambda}}}^{Correlación\ canonica} .$$

Es una medida de la asociación entre cada función discriminante y la variable que hay que explicar. El cuadrado de la correlación canónica indica el porcentaje de la varianza total de la variable dependiente (SCT = Suma de cuadrados totales), que se explica por la función discriminante correspondiente, donde SCT = SCinterg + SCintrag.

✓ **Descriptivos:** Las opciones disponibles son: Medias (que incluye las desviaciones estándar), Anovas univariados y prueba M de Box, Medias. Muestra la media y desviación estándar totales y las medias y desviaciones estándar de grupo, para las variables independientes.

✓ **Anovas univariados:** Realiza un análisis de varianza de un factor sobre la igualdad de las medias de grupo para cada variable independiente.

✓ **M de Box:** Contraste de la hipótesis nula sobre la igualdad de las matrices de covarianza de los grupos (varianzas-covarianzas poblacionales). Para tamaños de muestras suficientemente grandes, un valor de p no significativo quiere decir que no hay suficiente evidencia de que las varianzas sean diferentes. Esta prueba es sensible a las desviaciones de la normalidad multivariada. Uno de los supuestos del análisis discriminante es que todos los grupos proceden de la misma población y, más concretamente, que las matrices de varianzas-

covarianzas poblacionales correspondientes a cada grupo son iguales entre sí. El estadístico M de Box toma la forma:

$$M = (n-g)log|S| \sum_{j=1}^{g} (n_j - 1) log|S^j|,$$

donde S es matriz de varianzas – covarianzas combinadas, S^j es matriz de varianzas – covarianzas del j-ésimo grupo, n es número total de casos, n_j es número total de casos en el j-ésimo grupo, g es número de grupos.

✓ **Distancia de Mahalanobis**. Medida de cuánto difieren del promedio para todos los casos los valores en las variables independientes de un caso dado. Una distancia de Mahalanobis grande identifica un caso que tenga valores extremos en una o más de las variables independientes. Es la medida de la distancia entre dos puntos en el espacio, definido por dos o más variables correlacionadas. La distancia multivariante entre dos grupos a y b se define como:

$$H_{ab}^2 = (n-g) \sum_{i=1}^{p} \sum_{j=1}^{p} w_{ij}^* \left(\bar{x}_i^{(a)} - \bar{x}_i^{(b)}\right)\left(\bar{x}_j^{(a)} - \bar{x}_j^{(b)}\right),$$

donde n es número de casos válidos, g es número de grupos, $\bar{x}_i^{(a)}$ es media del grupo a en la i-esima variable independiente, $\bar{x}_i^{(b)}$ es media del grupo b en la i-esima variable independiente, w_{ij}^* es elemento de la inversa de la matriz de varianzas – covarianzas intragrupos.

✓ **Menor razón F:** Método para la selección de variables en los análisis por pasos que se basa en maximizar la razón F, calculada a partir de la distancia de Mahalanobis entre los grupos. Se incorpora en cada paso la variable que maximiza la menor razón de F para las parejas de los grupos. El estadístico F utilizado es la distancia de Mahalanobis ponderada por el tamaño de los grupos.

$$\frac{(n-p-1)n_1 n_2}{p(n-2)(n_1+n_2)} H_{ab}^2,$$

donde n es el número total de casos válidos, n_1 y n_2 es el número de casos en los grupos 1 y 2, p es el número de variables independientes o discriminantes.

✓ **V de Rao:** Medida de las diferencias entre las medias de los grupos. También se denomina la traza de Lawley-Hotelling. En cada paso, se incluye la variable que maximiza el incremento de la V de Rao. El estadístico V de Rao es una transformación de la traza de Lawley-Hotelling que es directamente proporcional a la distancia entre los grupos. Al utilizar este criterio la variable que se incorpora al modelo es aquella que produce un mayor incremento en el valor de V, se define como:

$$V = (n_a - g)\sum_{i=1}^{p}\sum_{j=1}^{p} w_{ij}^{*}\sum_{a=1}^{g}\left(\bar{x}_i^{(a)} - \bar{x}_i\right)\left(\bar{x}_j^{(a)} - \bar{x}_j\right),$$

donde p es número de variables en el modelo, g es número de grupos, n_a es número de casos válidos en el grupo a, $\bar{x}_i^{(a)}$ es media del grupo a en la i-ésima variable, $\bar{x}_i$ es Media de todos los grupos en la i-esima variable, w_{ij}^{*} es elemento de la inversa de la matriz de varianzas – covarianzas intragrupos.

✓ **Chi cuadrada (χ^2):** Es un indicador estadístico para determinar si existe una asociación sistemática entre dos variables nominales. Para ello se estima la probabilidad de obtener un valor de Chi cuadrada, a partir de la sumatoria de las diferencias entre las frecuencias observadas y las frecuencias esperadas. Este es aplicable para analizar relaciones entre variables cualitativas a las que no puede calcularse la media ni la varianza; además, existen un conjunto de pruebas no paramétricas que buscan probar hipótesis que no especifican condiciones sobre los parámetros de la población de la que proviene la muestra.

Una característica importante de este estadístico es el número de grados de libertad gl asociado a éste. Es decir, $gl = (r-1)(c-1)$ donde r = es el número de filas y c = el número de columnas. En la hipótesis nula (H_o) se determina que no hay relación entre las dos variables, se rechaza únicamente cuando el valor calculado del estadístico de prueba es mayor que el valor crítico de la distribución Chi cuadrada, con el número apropiado de grados de libertad.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} .$$

Donde f_o es la frecuencia observada o real, f_e es la frecuencia esperada, $\sum$ es la sumatoria, χ^2 es el estadístico Chi cuadrada, n número total de datos en la tabla.

$$f_e = \frac{T_f * T_c}{n} .$$

Donde f_e es la frecuencia esperada en la posición determinada, T_f es el total de filas en esa posición, T_c es el total de columnas en esa posición, n número total de datos en la tabla.

✓ **Normalización de los datos de entrada:** dado que está considerando un problema en el que se pretende estimar el valor futuro de un activo, y los valores de entrada al modelo tienen magnitudes diferentes, esto puede generar problemas con la etapa de aprendizaje y entrenamiento de los modelos. Para evitar esta situación, fue necesario normalizar los datos de entrada, realizando una técnica de preprocesamiento de los datos, la cual busca que los datos de entrada y de salida quedaran en un intervalo [-1,+1]. Existen dos formas comunes para normalizar los datos. La primera normaliza el dato, a partir del valor máximo del conjunto de datos y del valor mínimo. El segundo, utiliza la técnica de normalización gaussiana.

$$x_n = 2\left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) - 1 , \quad x_n = \left(\frac{x - \mu}{\sigma}\right),$$

donde μ es el valor medio de los datos y σ es la desviación estándar de los datos.

✓ **Grupos combinados:** Crea un diagrama de dispersión, con todos los grupos, de los valores en las dos primeras funciones discriminantes. Si sólo hay una función, en su lugar se muestra un histograma.

✓ **Grupos separados:** Crea diagramas de dispersión, de los grupos por separado, para los valores en las dos primeras funciones discriminantes. Si sólo hay una función, en su lugar se muestra un histograma.

3. Resultados y análisis

3.1. Resultado análisis discriminante

Tabla 9. Resumen de procesamiento de casos de análisis

Casos sin ponderar		N	Porcentaje
Válido		184	100
Excluido	Códigos de grupo perdidos o fuera de rango	0	0
	Como mínimo, falta una variable discriminatoria	0	0
	Faltan ambos códigos de grupo, los perdidos o los que están fuera de rango y, como mínimo, una variable discriminatoria	0	0
	Total	0	0
Total		184	100

En la Tabla 9 se muestra el total de encuestados en el grado octavo o casos analizados e indica que todos fueron válidos para el estudio; los casos fueron 184 y validos el 100%.

Se estudiaron varias alternativas de análisis que se anuncian en esta sección. Se indican, en cada una de ellas, las dificultades encontradas con respecto al porcentaje de clasificación errada. Finalmente se explica, en forma detallada, los resultados del AD definitivo.

3.1.1. Análisis para 4 grupos 14 variables predictoras.

Hay una variable grupo la cual permite clasificar los datos y definir cuatro grupos de estudiantes, además se toma 14 variables predictoras para este análisis;

G1 (rendimiento Bajo nota promedio $1{,}0 < x \leq 5{,}9$): notas desde 1,0 hasta 5,9.

G2 (rendimiento Básico nota promedio $6{,}0 \leq x \leq 7{,}9$): notas desde 6 hasta 7,9.

G3 (rendimiento Alto nota promedio $8{,}0 \leq x \leq 8{,}9$): notas desde 8 hasta 8,9.

G4 (rendimiento Superior nota promedio $9{,}0 \leq x \leq 10{,}0$): notas desde 9 hasta 10,0.

G1, G2, G3 y G4 indica los grupos de pertenencia.

Para este análisis en el procesamiento de casos se tomaron en cuenta 184 casos válidos y en las siguientes tablas se muestra la variación de los datos de la medias y la desviación estándar por variable en cada grupo y en el total. En las Tablas 10 y 11 se muestra un paralelo (Realizado por el programa spss y organizado por el autor) de los cuatro grupos comparando las medias

aritméticas y las desviaciones estándar en una correlación con cada una de las variables, llamada estadísticas de grupo.

Tabla 10. Media de la variable en cada grupo y en el total ponderado.

	Medias				
N° de estudiantes.	**4**	**56**	**20**	**4**	
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**
Perdidas	1,00	2,21	3,00	3,00	2,28
Tiempodedif	2,75	2,57	3,20	2,75	2,65
Entregatareas	3,25	3,44	4,00	4,50	3,52
Padrespend	2,25	3,68	3,90	3,75	3,67
Vivepadres	3,50	3,40	3,30	3,50	3,39
Ambientefam	4,25	4,15	4,35	4,00	4,17
Educacionrec	4,50	4,18	4,20	4,50	4,20
Tegustacol	3,75	4,04	4,25	3,75	4,05
Haspenfu	5,00	4,40	4,30	5,00	4,42
Vicbullyng	1,25	1,75	1,90	1,00	1,74
Nivelespa	2,75	2,79	2,90	2,75	2,80
Cartracasa	3,25	2,00	1,90	1,75	2,01
Respetocodoc	4,00	3,76	4,30	4,00	3,83
Pactoconv	2,75	3,26	3,55	3,25	3,28

En la Tabla 10 se muestra que los mayores datos de las medias están entre 3,9 y 5,0 pero se debe tener en cuenta cuales fueron los valores de los datos codificados y recodificados para cada variable.

Tabla 11. Desviación estándar de la variable en cada grupo y en el total ponderado.

Desviación estándar					
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**
Perdidas	0,000	0,751	0,000	0,000	0,766
Tiempodedif	1,258	0,931	1,196	0,500	0,975
Entregatareas	1,500	0,829	1,214	0,577	0,911
Padrespend	1,258	1,249	1,294	0,957	1,260
Vivepadres	1,000	0,679	0,733	0,577	0,685
Ambientefam	0,957	1,168	1,268	0,816	1,164
Educacionrec	1,000	1,075	1,281	0,577	1,084
Tegustacol	1,893	1,124	1,164	0,500	1,132
Haspenfu	0,000	1,076	1,302	0,000	1,083
Vicbullyng	0,500	1,298	1,518	0,000	1,300
Nivelespa	0,500	0,598	0,718	0,500	0,605
Cartracasa	2,062	1,229	1,373	0,500	1,259
Respetocodoc	1,414	1,246	0,923	0,816	1,216
Pactoconv	2,062	1,106	1,099	0,957	1,123

En la Tabla 11 vemos la desviación estándar en cada variable, como se está comportando el sesgo de los datos, donde 0,00 indica valores iguales en la variable para ese grupo determinado; un valor distinto indica que hay datos diferentes y entre más grande la desviación mayor diferencia entre los datos de cada variable.

Estos dos estadísticos permiten hacernos una idea de cómo están distribuidos los datos en cada variable, si tienen valores altos o bajos y la diferencia que puede haber entre ellos.

Tabla 12. Prueba de igualdad de medias de grupos.

Prueba de igualdad de medias de grupos					
Variables	Lambda de Wilks	F	gl1	gl2	Sig.
Perdidas	0,815	13,634	3	180	0,000
Tiempodedif	0,959	2,557	3	180	0,057
Entregatareas	0,936	4,084	3	180	0,008
Padrespend	0,968	1,954	3	180	0,123
Vivepadres	0,997	0,187	3	180	0,905
Ambientefam	0,996	0,211	3	180	0,889
Educacionrec	0,996	0,219	3	180	0,883
Tegustacol	0,993	0,397	3	180	0,756
Haspenfu	0,986	0,856	3	180	0,465
Vicbullyng	0,988	0,722	3	180	0,540
Nivelespa	0,997	0,198	3	180	0,897
Cartracasa	0,977	1,414	3	180	0,240
Respetocodoc	0,980	1,244	3	180	0,295
Pactoconv	0,988	0,702	3	180	0,552

En la Tabla 12 se muestra que de las 14 variables hay 2 variables (Perdidas, Entregatareas) que pueden ser consideradas porque sus valores reflejan un alto poder discriminante; pero se puede considerar la variable Tiempodedif por su importancia en este estudio, aunque se debe ser cauteloso por el valor de F.

La Lambda de Wilks, estadístico que mide el poder discriminante de un conjunto de variables y además, toma valores entre 0 y 1, siendo más discriminante cuando se acerca más a cero. Este estadístico aunque en la Tabla 12 es alto para las 3 variables consideradas, muestra los valores más bajos en: Pérdidas 0.815, Tiempodedif 0.959 y Entregatareas 0.936.

En el estadístico F, la variable pasa a formar parte de la función discriminante si el valor es mayor de 3,84 como valor de entrada y es expulsada de la función si el valor del estadístico es menor de 2,71 (F > 2,71 para aceptar) como valor de salida. Para este análisis podemos ver que de las 3 variables seleccionadas inicialmente solo hay 2 que cumplen con el criterio: **Perdidas**

13.634, **Entregatareas** 4.084, aunque se asume incluir la variable **Tiempodedif** 2.557 ya que su valor F está muy cercano de 2.71 con un nivel crítico de 0.057 muy cercano de 0.05, que para las otras variables es 0.00 y 0.08 respectivamente; esto se considera dada la importancia del tiempo dedicado a estudiar por fuera de clases y lo necesario que es para obtener un buen rendimiento académico.

3.1.1.1. Interpretación de la covarianza

Permite medir la asociación lineal que hay entre dos variables aleatorias o dicho de otra forma el grado de variación conjunta entre dos variables aleatorias respecto a sus medias. Si el valor es negativo hay asociación lineal negativa y si es positivo hay asociación lineal positiva.

Si $S_{xy} > 0$ hay dependencia directa (positiva), es decir, a grandes valores de x corresponden grandes valores de y.

Si $S_{xy} = 0$ una covarianza 0 se interpreta como la no existencia de una relación lineal entre las dos variables estudiadas.

Si $S_{xy} < 0$ hay dependencia inversa o negativa, es decir, a grandes valores de x corresponden pequeños valores de y.

Tabla 13. Matriz de covarianza dentro de grupos combinados, con 180 grados de libertad.

Matrices dentro de grupos combinados[a]															
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Convives	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Covarianza	Perdidas	**0,486**	0,160	0,116	0,024	0,029	0,040	0,074	0,038	-0,022	-0,011	0,048	-0,011	0,021	0,021
	Tiempodedif	0,160	**0,927**	0,185	0,122	-0,009	-0,102	-0,038	0,028	-0,012	0,122	0,023	-0,020	-0,086	0,155
	Entregatareas	0,116	0,185	**0,790**	0,113	0,028	0,225	0,306	0,184	0,384	-0,211	-0,028	-0,221	0,104	0,078
	Padrespend	0,024	0,122	0,113	**1,563**	0,025	0,310	0,208	0,125	0,321	-0,139	0,114	-0,115	0,391	0,191
	Convives	0,029	-0,009	0,028	0,025	**0,475**	-0,015	-0,102	-0,094	-0,060	0,014	0,063	-0,008	0,013	-0,001
	Ambientefam	0,040	-0,102	0,225	0,310	-0,015	**1,372**	0,450	0,431	0,615	-0,371	-0,041	-0,292	0,347	0,164
	Educacionrec	0,074	-0,038	0,306	0,208	-0,102	0,450	**1,190**	0,688	0,664	-0,662	-0,021	-0,237	0,476	0,320
	Tegustacol	0,038	0,028	0,184	0,125	-0,094	0,431	0,688	**1,295**	0,512	-0,393	-0,015	0,014	0,400	0,265
	Haspenfu	-0,022	-0,012	0,384	0,321	-0,060	0,615	0,664	0,512	**1,176**	-0,559	-0,053	-0,269	0,403	0,214
	Vicbullyng	-0,011	0,122	-0,211	-0,139	0,014	-0,371	-0,662	-0,393	-0,559	**1,699**	-0,016	0,225	-0,699	-0,170
	Nivelespa	0,048	0,023	-0,028	0,114	0,063	-0,041	-0,021	-0,015	-0,053	-0,016	**0,371**	-0,029	0,077	0,024
	Cartracasa	-0,011	-0,020	-0,221	-0,115	-0,008	-0,292	-0,237	0,014	-0,269	0,225	-0,029	**1,574**	-0,169	-0,024
	Respetocodoc	0,021	-0,086	0,104	0,391	0,013	0,347	0,476	0,400	0,403	-0,699	0,077	-0,169	**1,472**	0,452
	Pactoconv	0,021	0,155	0,078	0,191	-0,001	0,164	0,320	0,265	0,214	-0,170	0,024	-0,024	0,452	**1,268**

Para el caso de este estudio en el Tabla 13, si analizamos los valores de covarianza para las variables seleccionadas vemos que hay una dependencia directamente positiva en algunas y negativas en otras, pero en este caso los valores no son tan grandes. Se nota valores de covarianza positivos en las variables **Pérdidas, Entregatareas y Tiempodedif**, indicando que hay dependencia directa entre los valores de cada par de variables. Es de aclarar que en la tabla hay variables correlacionadas con mejor covarianza.

3.1.1.2. Interpretación de la correlación

Coeficiente de correlación es una medida numérica que permite medir el grado de asociación lineal entre dos variables cuantitativas.

- Si no existe ninguna relación entre las dos variables, la correlación se aproxima a 0.
- Si la correlación está cerca de 1 o −1, entonces hay una relación aproximadamente lineal.
- Si la correlación es de 1 o −1, entonces hay una relación lineal perfecta. Los valores dentro de la matriz de variables para la correlación son:

Tabla 14. Valores de la correlación (video youtube).

INTERVALO DE VALORES	TIPO DE CORRELACIÓN
± 0,96 - ± 1,0	Perfecta
± 0,85 - ± 0,95	Fuerte
± 0,70 - ± 0,84	Significativa
± 0,50 - ± 0,69	Moderada
± 0,20 - ± 0,49	Débil
± 0,10 - ± 0,19	Muy Débil
± 0,09 - ± 0,0	Nula

En el Tabla 15, si analizamos los valores de correlación para las variables seleccionadas, vemos que entre los datos que se muestran hay correlaciones débiles en algunas variables y nulas en otras, dado que la asociación lineal es baja, mostrándonos que los datos están dispersos.

Se nota valores de correlación positiva en las variables **Pérdidas, Entregatareas y Tiempodedif**, aunque no son las mejores correlaciones de todo la Tabla 15, pero tienen un alto grado de influencia en el estudio que se hace.

Tabla 15. Matriz de correlación dentro de grupos combinados, con 180 grados de libertad.

Matrices dentro de grupos combinados[a]														
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Convives	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Correlación	Perdidas	**1,000**	0,238	0,187	0,027	0,061	0,050	0,097	0,047	-0,029	-0,012	0,112	-0,013	0,025	0,027
	Tiempodedif	0,238	**1,000**	0,217	0,101	-0,013	-0,090	-0,036	0,026	-0,011	0,097	0,039	-0,017	-0,074	0,143
	Entregatareas	0,187	0,217	**1,000**	0,102	0,046	0,217	0,316	0,182	0,398	-0,182	-0,052	-0,198	0,097	0,078
	Padrespend	0,027	0,101	0,102	**1,000**	0,029	0,212	0,152	0,088	0,237	-0,085	0,150	-0,073	0,258	0,136
	Convives	0,061	-0,013	0,046	0,029	**1,000**	-0,019	-0,135	-0,120	-0,081	0,016	0,150	-0,009	0,015	-0,001
	Ambientefam	0,050	-0,090	0,217	0,212	-0,019	**1,000**	0,352	0,324	0,484	-0,243	-0,057	-0,199	0,244	0,124
	Educacionrec	0,097	-0,036	0,316	0,152	-0,135	0,352	**1,000**	0,555	0,561	-0,465	-0,032	-0,173	0,359	0,261
	Tegustacol	0,047	0,026	0,182	0,088	-0,120	0,324	0,555	**1,000**	0,415	-0,265	-0,022	0,010	0,290	0,207
	Haspenfu	-0,029	-0,011	0,398	0,237	-0,081	0,484	0,561	0,415	**1,000**	-0,396	-0,080	-0,198	0,306	0,175
	Vicbullyng	-0,012	0,097	-0,182	-0,085	0,016	-0,243	-0,465	-0,265	-0,396	**1,000**	-0,021	0,138	-0,442	-0,116
	Nivelespa	0,112	0,039	-0,052	0,150	0,150	-0,057	-0,032	-0,022	-0,080	-0,021	**1,000**	-0,038	0,104	0,035
	Cartracasa	-0,013	-0,017	-0,198	-0,073	-0,009	-0,199	-0,173	0,010	-0,198	0,138	-0,038	**1,000**	-0,111	-0,017
	Respetocodoc	0,025	-0,074	0,097	0,258	0,015	0,244	0,359	0,290	0,306	-0,442	0,104	-0,111	**1,000**	0,331
	Pactoconv	0,027	0,143	0,078	0,136	-0,001	0,124	0,261	0,207	0,175	-0,116	0,035	-0,017	0,331	**1,000**

3.1.1.3. La correlación canónica

Es la correlación entre la combinación lineal de las variables independientes (la función discriminante) y una combinación lineal de variables indicador (unos y ceros) que recogen la pertenencia de los sujetos a los grupos. En el caso de dos grupos, la correlación canónica es la correlación simple entre las puntuaciones discriminantes y una variable con códigos 1 y 0 según cada caso pertenezca a un grupo o a otro; por tal motivo una correlación canónica alta indica que las variables discriminantes permiten diferenciar entre los grupos[17].

Los Autovalores que se han obtenido en la Tabla 16, reflejan datos bastante cercanos a 0 y las correlaciones canónicas reflejan valores moderados, por lo que se supone que las variables discriminantes utilizadas no permiten distinguir demasiado bien entre los 4 grupos; además, cada una de ellas con un porcentaje de varianza de discriminación equivalente a 69.90%,

[17] Análisis discriminante: El procedimiento Discriminante. Capítulo 23

20.4% y 9.7% respectivamente y el porcentaje acumulado explicado por cada función discriminante es de 69.90%, 90.30% y 100%.

De otro lado, se logra identificar que las variables resultan significativas en la prueba de análisis de varianza entre los cuatro grupos escogidos. Este es un indicador de que, en al menos dos de los grupos, se produjeron diferencias significativas respecto a las variables tenidas en cuenta en la discriminación.

Tabla 16. Resumen de funciones discriminantes canónicas.

Se utilizaron las primeras 3 funciones discriminantes canónicas en el análisis.

Autovalores				
Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	0,330[a]	69,90	69,90	0,50
2	0,096[a]	20,40	90,30	0,30
3	0,046[a]	9,70	100,00	0,21

La gran ventaja diagnóstica del estadístico Lambda es que, puesto que se basa en las matrices de varianzas-covarianzas, puede calcularse antes de obtener las funciones discriminantes.

En la Tabla 17, el valor de Lambda es moderadamente medio-alto (0.655) en la primera función, lo cual significa que existe buen solapamiento entre los grupos. Sin embargo, el valor transformado de lambda (Chi-cuadrado = 73.503) tiene asociado, con 42 grados de libertad, un nivel crítico (Sig.) de 0.002, por lo que podemos rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales en las variables discriminantes en cuestión.

En la segunda función el valor de lambda es alto (0.872), lo cual significa que existe un gran solapamiento entre los grupos y que el valor transformado de lambda (Chi-cuadrado = 23.820) tiene asociado, con 26 grados de libertad, un nivel crítico (Sig.) de 0.586, por lo que no se rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales pero con estos valores se indica que pueden parecerse un poco en las variables discriminantes en cuestión.

Para la tercera función el valor de lambda es muy alto (0.956), lo cual significa que existe un alto solapamiento, bastante fuerte entre los grupos y que el valor transformado de lambda (Chi-cuadrado = 7.815) tiene asociado, con 12 grados de libertad, un nivel crítico (Sig.) de 0.799, por lo que no se rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales, esto debido a que el valor Lambda de Wilks es tan cercano que parece a 1 (uno) y esto indica medias muy parecidas, casi que iguales en las variables discriminantes en cuestión.

Tabla 17. Lambda de Wilks

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a 3	0,655	73,503	42	0,002
2 a 3	0,872	23,820	26	0,586
3	0,956	7,815	12	0,799

3.1.1.4. La matriz de estructuras.

Es importante y de interés ya que permite conocer cómo se relaciona cada variable independiente con la función discriminante. Conocer estas relaciones puede ayudar a interpretar mejor la función discriminante.

Existe Correlaciones dentro de grupos combinados entre las variables discriminantes y las funciones discriminantes canónicas estandarizadas. En la Tabla 18 Las variables están ordenadas por el tamaño absoluto de la correlación dentro de la función y el mayor valor, este orden puede ser distinto del orden en el que aparecen en otras tablas y del orden en que han sido incluidas en el análisis. Se indica en cada fila la mayor correlación absoluta (posición del asterisco) entre cada variable y cualquier función discriminante.

En la Tabla 18 podemos apreciar que hay tres funciones, las cuales están asociadas y correlacionadas de la siguiente forma:

- En la función 1 las variables que tienen mayor correlación son Pérdidas 0.823, Entregatareas 0.362, Padrespend 0.269 y Tiempodedif 0.242, dando mayor importancia a la disciplina académica y la responsabilidad de los padres.
- En la función 2 las variables que tienen mayor correlación son: Tiempodedif 0.458, Entregatareas 0.415, Respetocodoc 0.348 y Padrespend -0.299, dando mayor importancia a la disciplina académica, el respeto hacia la comunidad educativa y la responsabilidad de los padres que muestra una menor importancia en la función.
- En la función 3 las variables que tienen mayor correlación son: Vicbullyng 0.425 Entregatareas -0.424, Haspenfu -0.393 y Tegustacol 0.261, dando mayor importancia la convivencia y el gusto por la institución, por otra parte muestra una menor importancia el cumplimiento por las actividades y la proyección a futuro. Tener en cuenta que la mejor correlación se acerca a +1 y el signo positivo muestra la mayor importancia en la función.

Tabla 18. La matriz de estructuras

	Función		
	1	2	3
Perdidas	0,823*	0,164	-0,15
Pactoconv	0,181*	-0,012	0,137
Tiempodedif	0,242	0,458*	0,254
Respetocodoc	0,158	0,348*	0,129
Padrespend	0,269	-0,299*	0,027
Cartracasa	-0,214	0,285*	0,115
Vicbullyng	0,076	-0,136	0,425*
Entregatareas	0,362	0,415	-0,424*
Haspenfu	-0,098	0,203	-0,393*
Tegustacol	0,102	-0,013	0,261*
Ambientefam	0,047	0,115	0,182*
Convives	-0,073	-0,011	-0,171*
Educacionrec	-0,025	0,148	-0,170*
Nivelespa	0,077	0,064	0,143*

3.1.1.5. Resultados de clasificación para 4 grupos.

En el Tabla 19 se muestra que el 60.3% de casos agrupados fueron clasificados correctamente; los otros 39,7% de los casos fueron clasificados erróneamente. Debe destacarse que los grupos Básico, Alto y Superior presentan porcentajes muy elevados de clasificación errónea (42.3%, 30% y 25%, respectivamente); además, el grupos Bajo tuvo un muy buen porcentaje de buena clasificación con el 100%, aunque se debe tener en cuenta que para este grupo solo se presentaron 4 casos.

Tabla 19. Resultados de clasificación

60,3% de casos originales agrupados clasificados correctamente.

Resultados de clasificación [a]								
		Nivel	Pertenencia a grupos pronosticada				Total	
			Bajo	Básico	Alto	Superior		
Original	Recuento	Bajo	4	0	0	0	4	184
		Basico	14	90	29	23	156	
		Alto	0	2	14	4	20	
		Superior	0	0	1	3	4	
	%	Bajo	100%	0,0%	0,0%	0,0%	100%	
		Basico	9,0%	57,7%	18,6%	14,7%	100%	
		Alto	0,0%	10%	70%	20%	100%	
		Superior	0,0%	0,0%	25%	75%	100%	

Como el análisis y la clasificación que muestra la Tabla 19 se cree que es baja, se plantea otra opción o nueva alternativa en la cual se definen 3 grupos de Bajos, Medio y Alto rendimiento.

3.1.2. Análisis para 3 grupos y 14 variables predictoras.

Se usa nuevamente la variable grupo (Nota promedio: Nivel de competencia) la cual permite clasificar los datos y definir tres grupos de estudiantes, además se toma 14 variables predictoras para este análisis:

G1 (rendimiento Bajo nota promedio $1{,}0 < x \leq 5{,}9$): notas desde 1,0 hasta 5,9.

G2 (rendimiento Medio nota promedio $6{,}0 \leq x \leq 8{,}0$): notas desde 6,0 hasta 8,0.

G3 (rendimiento Alto nota promedio $8{,}1 \leq x \leq 10{,}0$): notas desde 8,1 hasta 10,0.

G1, G2 y G3 indica los grupos de pertenencia.

Para este análisis el procesamiento de casos es similar al anterior, se tomaron en cuenta 184 casos válidos y en las Tablas 20 y 21 se muestra la variación de los datos de la medias y la desviación estándar por variable para tres (3) grupos y en el total, llamada estadísticas de grupo.

Tabla 20. Media de las variables en cada grupo y en el total ponderado.

MEDIAS				
CASOS VÁLIDOS	4	158	22	184
VARIABLES	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	1,000	2,220	3,000	2,280
Tiempodedif	2,750	2,560	3,230	2,650
Entregatareas	3,250	3,460	4,000	3,520
Padrespend	2,250	3,650	4,090	3,670
Convives	3,500	3,400	3,320	3,390
Ambientefam	4,250	4,150	4,320	4,170
Educacionrec	4,500	4,190	4,180	4,200
Tegustacol	3,750	4,050	4,090	4,050
Haspenfu	5,000	4,410	4,360	4,420
Vicbullyng	1,250	1,740	1,820	1,740
Nivelespa	2,750	2,800	2,860	2,800
Cartracasa	3,250	2,010	1,770	2,010
Respetocodoc	4,000	3,770	4,230	3,830
Pactoconv	2,750	3,270	3,450	3,280

La Tabla 20 muestra que los mayores datos de las medias están entre 3,5 y 5,0 pero se debe tener en cuenta cuales fueron los valores de los datos codificados y recodificados para cada variable; cada valor es dado por el promedio de los datos de la variable en cada grupo.

Tabla 21. Desviación estándar de la variable en cada grupo y en el total ponderado.

DESVIACIÓN ESTANDAR				
CASOS VÁLIDOS	4	158	22	184
VARIABLES	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	0,000	0,752	0,000	0,766
Tiempodedif	1,258	0,934	1,066	0,975
Entregatareas	1,500	0,842	1,155	0,911
Padrespend	1,258	1,267	1,019	1,260
Convives	1,000	0,677	0,716	0,685
Ambientefam	0,957	1,167	1,211	1,164
Educacionrec	1,000	1,072	1,220	1,084
Tegustacol	1,893	1,122	1,109	1,132
Haspenfu	0,000	1,072	1,255	1,083
Vicbullyng	0,500	1,293	1,468	1,300
Nivelespa	0,500	0,595	0,710	0,605
Cartracasa	2,062	1,247	1,110	1,259
Respetocodoc	1,414	1,242	0,922	1,216
Pactoconv	2,062	1,108	1,057	1,123

El Tabla 21 contiene la desviación estándar en cada variable y como se está comportando el sesgo de los datos con relación a la media en cada variable y grupo, entre más grande la desviación mayor diferencia entre los datos de cada variable en el grupo.

Tabla 22. Prueba de igualdad de medias de grupos.

Prueba de igualdad de medias de grupos					
Variables	Lambda de Wilks	F	gl1	gl2	Sig.
Perdidas	0,826	19,002	2	181	0,000
Tiempodedif	0,951	4,680	2	181	0,010
Entregatareas	0,961	3,651	2	181	0,028
Padrespend	0,959	3,905	2	181	0,022
Vivepadres	0,998	0,183	2	181	0,833
Ambientefam	0,998	0,221	2	181	0,802
Educacionrec	0,998	0,160	2	181	0,852
Tegustacol	0,998	0,153	2	181	0,858
Haspenfu	0,993	0,605	2	181	0,547
Vicbullyng	0,996	0,321	2	181	0,726
Nivelespa	0,999	0,131	2	181	0,877
Cartracasa	0,975	2,367	2	181	0,097
Respetocodoc	0,984	1,440	2	181	0,240
Pactoconv	0,992	0,721	2	181	0,488

La Tabla 22 muestra las 14 variables de las cuales hay 4 variables que pueden ser consideradas (**Perdidas, Tiempodedif, Entregatareas, Padrespend**) porque sus valores reflejan poder F discriminante.

La Lambda de Wilks, este estadístico aunque en la Tabla 22 es alto para las 4 variables consideradas, muestra los valores más bajos en: **Pérdidas** 0.826, **Tiempodedif** 0.951, **Entregatareas** 0.961 y **Padrespend** 0.959 o sea débil poder discriminante ya que son valores bastante distanciados de cero (0).

En el estadístico F, Para este caso se puede afirmar que de las 4 variables seleccionadas en las 4 se cumplen con el criterio (3,84 como valor de entrada y 2,71 como valor de salida): **Perdidas** 19.002, **Tiempodedif** 4.680, **Entregatareas** 3.651 y **Padrespend** 3,905 con un niveles críticos de 0.000, 0.010, 0.028 y 0.022 respectivamente, menores que el nivel crítico de 0.05 por lo cual se puede rechazar la hipótesis de medias iguales.

Tabla 23. Matriz de covarianza dentro de grupos combinados, con 181 grados de libertad.

Matrices dentro de grupos combinados[a]															
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Vivepadres	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Covarianza	Perdidas	**0,490**													
	Tiempodedif	0,154	**0,914**												
	Entregatareas	0,129	0,183	**0,807**											
	Padrespend	0,005	0,104	0,100	**1,538**										
	Vivepadres	0,030	-0,009	0,030	0,026	**0,473**									
	Ambientefam	0,039	-0,101	0,222	0,305	-0,016	**1,366**								
	Educacionrec	0,080	-0,035	0,313	0,208	-0,101	0,447	**1,186**							
	Tegustacol	0,046	0,038	0,185	0,126	-0,096	0,433	0,683	**1,294**						
	Haspenfu	-0,017	-0,013	0,392	0,320	-0,058	0,608	0,664	0,503	**1,179**					
	Vicbullyng	-0,018	0,123	-0,223	-0,139	0,012	-0,365	-0,663	-0,383	-0,567	**1,704**				
	Nivelespa	0,049	0,025	-0,027	0,113	0,062	-0,039	-0,021	-0,013	-0,054	-0,014	**0,369**			
	Cartracasa	-0,002	-0,010	-0,216	-0,106	-0,009	-0,287	-0,237	0,014	-0,271	0,229	-0,028	**1,561**		
	Respetocodoc	0,027	-0,085	0,111	0,379	0,012	0,347	0,476	0,406	0,400	-0,694	0,078	-0,162	**1,471**	
	Pactoconv	0,027	0,159	0,082	0,188	-0,002	0,165	0,319	0,269	0,210	-0,166	0,026	-0,022	0,456	**1,265**

La Tabla 23 indica los valores de covarianza para las variables seleccionadas, las cuales tienen dependencia directamente positiva en algunas y negativas en otras, además, se muestra que los valores no tan grandes.

Las variables correlacionadas **Pérdidas, Entregatareas, Tiempodedif y Padrespend** indican que hay dependencia directa positiva entre los valores de cada par de variables, es decir, las varianzas reflejan datos con valores parecidos en tamaño. Es de aclarar que en la tabla hay variables correlacionadas con covarianza más altas es decir que los valores son más altos o mayor diferencia.

Tabla 24. Matriz de correlación dentro de grupos combinados, con 181 grados de libertad.

Matrices dentro de grupos combinados[a]														
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Vivepadres	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Correlación	Perdidas	**1,000**													
	Tiempodedif	0,230	**1,000**												
	Entregatareas	0,205	0,213	**1,000**											
	Padrespend	0,005	0,087	0,090	**1,000**										
	Vivepadres	0,062	-0,013	0,048	0,031	**1,000**									
	Ambientefam	0,048	-0,090	0,211	0,210	-0,020	**1,000**								
	Educacionrec	0,105	-0,033	0,320	0,154	-0,135	0,351	**1,000**							
	Tegustacol	0,057	0,034	0,181	0,090	-0,122	0,326	0,551	**1,000**						
	Haspenfu	-0,022	-0,013	0,402	0,238	-0,077	0,479	0,562	0,407	**1,000**					
	Vicbullyng	-0,019	0,099	-0,190	-0,086	0,013	-0,239	-0,466	-0,258	-0,400	**1,000**				
	Nivelespa	0,115	0,042	-0,050	0,149	0,148	-0,056	-0,032	-0,019	-0,082	-0,018	**1,000**			
	Cartracasa	-0,003	-0,008	-0,193	-0,068	-0,011	-0,197	-0,175	0,010	-0,200	0,140	-0,037	**1,000**		
	Respetocodoc	0,032	-0,073	0,102	0,252	0,014	0,245	0,360	0,294	0,304	-0,439	0,106	-0,107	**1,000**	
	Pactoconv	0,035	0,148	0,081	0,134	-0,003	0,126	0,260	0,210	0,172	-0,113	0,037	-0,016	0,334	**1,000**

En la Tabla 24, si analizamos los valores de correlación para las variables seleccionadas, vemos que entre los datos que se muestran hay correlaciones débiles en algunas variables y nulas en otras, dado que la asociación lineal es baja, mostrando que los datos están dispersos.

Se nota valores de correlación positiva débil y muy débil en las variables (O sea que al incrementar la una también incrementa la otra) **Pérdidas, Entregatareas, Tiempodedif y**

Padrespend, aunque no son las mejores correlaciones de toda la tabla, pero tienen un alto grado de influencia para el estudio.

Tabla 25. Resumen de funciones discriminantes canónicas.

Se utilizan las primeras 2 funciones discriminantes canónicas en el análisis

Autovalores				
Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	0,334[a]	80,4	80,4	0,500
2	0,081[a]	19,6	100	0,274

Los Autovalores obtenidos en la Tabla 25, muestran datos bastante cercanos a 0; para la primera función el Autovalor es de 0.334 que mediante cálculo lleva a una correlación canónica moderada de 0.500. Para la segunda función el Autovalor es de 0.081, dato algo bajo que al cálculo da una correlación canónica de 0.274. Estos resultados permiten inferir que las variables discriminantes utilizadas no distinguen demasiado bien los grupos.

Cada una de las funciones en la Tabla 25 con un porcentaje de varianza de discriminación equivalente a 80.40% y 19.60% respectivamente y el porcentaje acumulado explicado por cada función discriminante de 80.40% y 100%.

De otro lado, se logra identificar que las variables resultan significativas en la prueba de análisis de varianza entre los tres grupos escogidos. Este es un indicador de que, en al menos uno de los grupos, se produjeron diferencias significativas respecto a las variables tenidas en cuenta en la discriminación.

Tabla 26. Lambda de Wilks

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a 2	0,693	63,896	28	0,000
2	0,925	13,655	13	0,399

En la Tabla 26, el valor de Lambda es moderadamente medio-alto (0.693) en la primera función, lo cual significa que existe buen solapamiento entre los grupos. Sin embargo, el valor transformado de Lambda (Chi-cuadrado = 63.896) tiene asociado, con 28 grados de libertad, un

nivel crítico (Sig.) de 0.000, por lo que podemos rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales en las variables discriminantes en cuestión.

En la segunda función el valor de Lambda es alto (0.925), lo cual significa que existe un gran solapamiento entre los grupos y que el valor transformado de Lambda (Chi-cuadrado = 13.655) tiene asociado, con 13 grados de libertad, un nivel crítico (Sig.) de 0.399, por lo que no se rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales, esto debido a que el valor Lambda de Wilks es tan cercano que parece a 1 (uno) y esto indica medias muy parecidas, casi que iguales en las variables discriminantes en cuestión.

Tabla 27. Matriz de estructuras

Matriz de estructuras		
	Función	
	1	2
Perdidas	0,788*	0,187
Padrespend	0,351*	-0,158
Cartracasa	-0,257*	0,226
Pactoconv	0,153*	-0,046
Vivepadres	-0,076*	-0,032
Nivelespa	0,061*	0,052
Tiempodedif	0,265	0,590*
Entregatareas	0,299	0,358*
Respetocodoc	0,130	0,355*
Haspenfu	-0,114	0,169*
Ambientefam	0,048	0,143*
Educacionrec	-0,053	0,101*
Vicbullyng	0,092	-0,096*
Tegustacol	0,062	-0,072*

En la Tabla 27 podemos apreciar que hay dos funciones, las cuales están asociadas y correlacionadas de la siguiente forma:

- En la función 1 las variables que tienen mayor correlación son Pérdidas 0.788, Padrespend 0.351, Entregatareas 0.299 y Tiempodedif 0.265, dando mayor importancia a la disciplina académica y la responsabilidad de los padres.

- En la función 2 las variables que tienen mayor correlación son: Tiempodedif 0.590, Entregatareas 0.358, Respetocodoc 0.355 y Cartracasa 0.226, dando mayor importancia a la disciplina académica, el respeto hacia la comunidad educativa y la saturación de trabajo no académico en la casa.

Tabla 28. Resultados de clasificación para 3 grupos

68,5% de casos agrupados originales clasificados correctamente.

Resultados de clasificación[a]							
		Nivel1	Pertenencia a grupos pronosticada			Total	
			Bajo	Medio	Alto		
Original	Recuento	Bajo	4	0	0	4	184
		Medio	14	102	42	158	
		Alto	0	2	20	22	
	%	Bajo	100%	0,0%	0,0%	100%	
		Medio	8,9%	64,6%	26,6%	100%	
		Alto	0,0%	9,1%	90,9%	100%	

En la Tabla 28 se muestra que el 68.5% de casos agrupados fueron clasificados correctamente; los otros 31,5% de los casos fueron clasificados erróneamente. Se destaca que el grupo Medio presenta un porcentaje muy elevados de clasificación errónea (35.4%); además, los grupos Bajo y Alto tuvieron un muy buen porcentaje de buena clasificación con 100% y 90.9% respectivamente.

Como el porcentaje de clasificados correctamente en los dos análisis anteriores fue bueno, pero no lo suficiente, porque no supero el 70%, se opta por hacer un tercer análisis donde se utilizan dos grupos.

3.1.3. Análisis para 2 grupos y 14 variables predictoras.

Para este caso variable grupo (Nota promedio: Nivel de competencia) la cual permite clasificar los datos y se definen dos grupos de estudiantes, además se toma 14 variables predictoras para este análisis:

G1 (rendimiento Bajo nota promedio $1,0 < X \leq 5,9$): notas desde 1,0 hasta 5,9.

G2 (rendimiento Aprueba nota promedio $6,0 \leq X \leq 10,0$): notas desde 6,0 hasta 10,0.

G1 y G2 indica los grupos de pertenencia

Para este análisis en el procesamiento de casos es similar al anterior, con 184 casos válidos, en los cuales se distinguen dos grupos de acuerdo a su nota promedio, los que aprueban (Aprueba) y los que no aprueban (Bajo), clasificados de acuerdo a las notas o calificaciones que tiene la institución para identificar los aprobados y reprobados; o sea que en este tercer análisis sólo se utilizan dos grupos separados por la nota límite 6,0.

En la Tabla 29 se muestra la variación de los datos de la medias y la desviación estándar por variable para dos (2) grupos y en el total, estadísticas de grupo.

Tabla 29. Media y Desviación estándar de la variable en cada grupo y en el total ponderado; organizó autor.

	MEDIAS			DESVIACIÓN ESTANDAR		
CASOS VÁLIDOS	4	180	184	4	180	184
VARIABLES	**Bajo**	**Aprobado**	**Total**	**Bajo**	**Aprobado**	**Total**
Perdidas	1,000	2,310	2,280	0,000	0,750	0,766
Tiempodedif	2,750	2,640	2,650	1,258	0,972	0,975
Entregatareas	3,250	3,530	3,520	1,500	0,900	0,911
Padrespend	2,250	3,710	3,670	1,258	1,245	1,260
Convives	3,500	3,390	3,390	1,000	0,680	0,685
Ambientefam	4,250	4,170	4,170	0,957	1,170	1,164
Educacionrec	4,500	4,190	4,200	1,000	1,087	1,084
Tegustacol	3,750	4,060	4,050	1,893	1,117	1,132
Haspenfu	5,000	4,410	4,420	0,000	1,092	1,083
Vicbullyng	1,250	1,750	1,740	0,500	1,311	1,300
Nivelespa	2,750	2,810	2,800	0,500	0,608	0,605
Cartracasa	3,250	1,980	2,010	2,062	1,230	1,259
Respetocodoc	4,000	3,820	3,830	1,414	1,215	1,216
Pactoconv	2,750	3,290	3,280	2,062	1,101	1,123

La Tabla 29 muestra que los mayores datos de las medias están entre 3,500 y 5,000 para los grupos, entre 3.520 y 4.420 para el total teniendo en cuenta los valores de los datos codificados y recodificados para cada variable. Cada valor es dado por el promedio de los datos de la variable en cada grupo. Por otro lado se muestra la desviación estándar en cada variable y como se está comportando el sesgo de los datos con relación a la media en cada variable y grupo.

Tabla 30. Prueba de igualdad de medias de grupos

Prueba de igualdad de medias de grupos					
	Lambda de Wilks	F	gl1	gl2	Sig.
Perdidas	0,937	12,172	1	182	0,001
Tiempodedif	1,000	0,046	1	182	0,831
Entregatareas	0,998	0,362	1	182	0,548
Padrespend	0,971	5,348	1	182	0,022
Vivepadres	0,999	0,103	1	182	0,749
Ambientefam	1,000	0,020	1	182	0,888
Educacionrec	0,998	0,321	1	182	0,572
Tegustacol	0,998	0,284	1	182	0,595
Haspenfu	0,994	1,179	1	182	0,279
Vicbullyng	0,997	0,577	1	182	0,448
Nivelespa	1,000	0,033	1	182	0,856
Cartracasa	0,978	4,028	1	182	0,046
Respetocodoc	1,000	0,083	1	182	0,773
Pactoconv	0,995	0,900	1	182	0,344

En la Tabla 30 se muestra que de las 14 variables hay 3 variables (**Perdidas**, **Padrespend**, y **Cartracasa**) en donde sus valores reflejan un alto poder discriminante.

La Lambda de Wilks, es un estadístico que indica valores muy altos para las 3 variables consideradas (**Perdidas** 0.937, **Padrespend** 0.971, y **Cartracasa** 0.978) por tal motivo, podríamos decir que no hay buena discriminación si tomamos en cuenta este estadístico ya que el intervalo es de 0 hasta 1, siendo más discriminante cuando se acerca más a cero.

Para el estadístico F, podemos ver que las 3 variables seleccionadas cumplen con el criterio: **Perdidas** 12.172, **Padrespend** 5.348 y **Cartracasa** 4.028 ya que la variable es expulsada de la función si su valor F es menor de 2,71 (o sea que $F > 2{,}71$ para aceptar) como valor de salida; por otro lado, en este caso el nivel crítico o de significancia es de 0.001, 0.022 y 0.046 respectivamente, siendo valores muy importantes o significativos para considerar mantener las variables, y se rechaza la hipótesis en cada caso.

De lo anterior podemos notar que no se le está dando importancia a algunas variables que determinan la responsabilidad académica, variables que son importantes para el buen rendimiento académico; además, su valor F es muy bajo.

La matriz de covarianza en la Tabla 31 indica el grado de asociación lineal que hay entre variables aleatorias o el grado de variación conjunta entre variables aleatorias respecto a sus medias y podemos ver que es positiva en las variables **Perdidas** y **Padrespend** pero es negativa en **Cartracasa**; esto indica que donde es positiva la covarianza los datos guardan relación directa en cuanto a su tamaño en las variables relacionadas y donde la covarianza es negativa los datos guardan relación inversa en cuanto a su tamaño en las variables relacionadas. Otras variables que son importantes para el rendimiento académico como **Tiempodedif** y **Entregatareas** presentan covarianza positiva. Esto se muestra en el Tabla 31:

Tabla 31. Matriz de covarianza dentro de grupos combinados

Matrices dentro de grupos combinados[a]															
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Vivepadres	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Covarianza	Perdidas	0,553													
	Tiempodedif	0,208	0,956												
	Entregatareas	0,173	0,220	0,833											
	Padrespend	0,041	0,134	0,125	1,550										
	Vivepadres	0,023	-0,014	0,025	0,023	0,471									
	Ambientefam	0,053	-0,088	0,230	0,311	-0,017	1,361								
	Educacionrec	0,079	-0,035	0,311	0,206	-0,100	0,444	1,179							
	Tegustacol	0,049	0,040	0,187	0,127	-0,096	0,432	0,679	1,287						
	Haspenfu	-0,020	-0,017	0,387	0,316	-0,057	0,603	0,661	0,500	1,172					
	Vicbullyng	-0,011	0,128	-0,217	-0,135	0,011	-0,361	-0,659	-0,380	-0,565	1,695				
	Nivelespa	0,054	0,029	-0,024	0,115	0,061	-0,038	-0,021	-0,013	-0,054	-0,014	0,368			
	Cartracasa	-0,022	-0,026	-0,229	-0,116	-0,007	-0,290	-0,236	0,013	-0,268	0,225	-0,029	1,559		
	Respetocodoc	0,066	-0,052	0,137	0,399	0,008	0,353	0,473	0,405	0,395	-0,687	0,081	-0,173	1,485	
	Pactoconv	0,043	0,172	0,092	0,195	-0,004	0,168	0,317	0,268	0,208	-0,163	0,027	-0,027	0,463	1,262

Tabla 32. Matriz de correlación dentro de grupos combinados

Matrices dentro de grupos combinados[a]															
		Perdidas	Tiempodedif	Entregatareas	Padrespend	Vivepadres	Ambientefam	Educacionrec	Tegustacol	Haspenfu	Vicbullyng	Nivelespa	Cartracasa	Respetocodoc	Pactoconv
Correlación	Perdidas	1,000													
	Tiempodedif	0,287	1,000												
	Entregatareas	0,255	0,246	1,000											
	Padrespend	0,044	0,110	0,110	1,000										
	Vivepadres	0,045	-0,021	0,040	0,026	1,000									
	Ambientefam	0,061	-0,077	0,216	0,214	-0,022	1,000								
	Educacionrec	0,098	-0,033	0,314	0,152	-0,134	0,351	1,000							
	Tegustacol	0,058	0,036	0,180	0,090	-0,123	0,326	0,551	1,000						
	Haspenfu	-0,025	-0,016	0,392	0,234	-0,077	0,478	0,562	0,407	1,000					
	Vicbullyng	-0,011	0,100	-0,183	-0,083	0,012	-0,238	-0,466	-0,258	-0,400	1,000				
	Nivelespa	0,121	0,049	-0,042	0,152	0,147	-0,054	-0,032	-0,018	-0,082	-0,017	1,000			
	Cartracasa	-0,024	-0,022	-0,201	-0,075	-0,009	-0,199	-0,174	0,009	-0,198	0,139	-0,039	1,000		
	Respetocodoc	0,073	-0,043	0,123	0,263	0,009	0,249	0,357	0,293	0,300	-0,433	0,110	-0,114	1,000	
	Pactoconv	0,051	0,156	0,090	0,140	-0,005	0,128	0,260	0,211	0,171	-0,112	0,039	-0,019	0,338	1,000

La matriz de correlación, Tabla 32 indica el grado de incremento que hay entre los datos de las variables aleatorias correlacionadas, además, positiva en algunos y negativas en otras. Podemos ver el grado de dispersión.

En la Tabla 32 la correlación, es positiva en las variables **Perdidas** y **Padrespend (0.041)** pero es negativa en **Perdidas** y **Cartracasa (-0.022)**; esto indica que donde es positiva la correlación, las variables son directamente proporcionales, pero como el datos es pequeño indica una alta dispersión de puntos en el espacio.

Otras variables que son importantes para el rendimiento académico como **Tiempodedif** y **Entregatareas** presentan correlación positiva y valor más alto.

Tabla 33. Resumen de funciones discriminantes canónicas para dos grupos

Se utiliza la primera 1 función discriminante canónica en el análisis.

Autovalores				
Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	0,188	100	100	0,398

La correlación canónica que muestra en la Tabla 33 es un valor de 0.398 lo cual es un dato relativamente bajo ya que el valor se puede mover entre un intervalo de 0 y 1 e indica que a correlaciones canónicas altas las variables discriminantes permiten diferenciar mejor entre los grupos. Por otro lado el Autovalor que se han obtenido, reflejan un dato bastante cercanos a cero (0) por lo que se supone que las variables discriminantes utilizadas no permiten distinguir demasiado bien entre los 2 grupos; además, se muestra un porcentaje de varianza de discriminación equivalente a 100% e igualmente el acumulado explicado para la función discriminante es del 100%.

Tabla 34. Lambda de Wilks para dos grupos.

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	0,841	30,213	14	0,007

En la Tabla 34, el valor de Lambda es alto (0.841) en la función, lo cual significa que existe buen solapamiento entre los grupos y el valor transformado de Lambda (Chi-cuadrado = 30.213) tiene asociado, con 14 grados de libertad, un nivel crítico (Sig.) de 0.007, por lo que podemos aceptar la hipótesis nula de que los grupos comparados tienen promedios relativamente iguales en las variables discriminantes en cuestión; porque, estos valores indican que los grupos pueden parecerse un poco en las variables discriminantes en cuestión.

Tabla 35. Coeficientes de función discriminante canónica estandarizados y Matriz de estructuras.

	Coeficientes de función discriminante canónica estandarizados	Matriz de estructuras
Función	**1**	**1**
Perdidas	-0,704	-0,596
Tiempodedif	0,418	-0,395
Entregatareas	-0,071	0,343
Padrespend	-0,542	0,185
Vivepadres	0,121	-0,162
Ambientefam	0,166	-0,13
Educacionrec	0,353	-0,103
Tegustacol	-0,373	0,097
Haspenfu	0,286	-0,091
Vicbullyng	0,002	0,055
Nivelespa	0,133	0,049
Cartracasa	0,465	0,036
Respetocodoc	0,257	-0,031
Pactoconv	-0,275	0,024

La matriz de estructuras, Tabla 35, permite conocer cómo se relaciona cada variable independiente con la función discriminante. Las variables están ordenadas por el tamaño absoluto de la correlación dentro de la función, En la tabla podemos apreciar que hay una función, la cual está asociada y correlacionada de la siguiente forma: Las variables que tienen mayor correlación positiva o negativa son **Pérdidas** (-0.596), **Tiempodedif** (-0.395), **Entregatareas** (0.343) y **Padrespend** (0.185); esto indica la relación bruta entre cada variable y la función discriminante. Este orden puede ser distinto del orden en el que aparecen en otras tablas y del orden en que han sido incluidas en el análisis.

La Tabla 35 en el Coeficientes de función discriminante canónica estandarizados muestra que variables tienen mayor importancia en la función esto dependiendo su valor y su signo por correlación. Para este caso las variables **Perdidas** (-0.704) y **Padrespend** (-0.542) correlacionan alto con la función discriminante, pero por su valor muestran ser variables poco valorativas en la función, esto puede deberse a el tipo de variables con la cual se han relacionado. Las variables **Tiempodedif** (0.418) y **Cartracasa** (0.465) por su valor muestran que correlacionan moderadamente medio siendo positivas, por lo cual son variables importantes en la valoración de la función discriminante.

Tabla 36. Resultados de clasificación, dos grupos

91,3% de casos agrupados correctamente clasificados.

		Nivel1	Pertenencia a grupos pronosticada		Total	
			Bajo	Aprueba		
Original	Recuento	Bajo	4	0	4	184
		Aprueba	16	164	180	
	%	Bajo	100%	0.0%	100%	
		Aprueba	8,9%	91,1%	100%	

El Tabla 36 muestra el porcentaje de clasificación o casos originales clasificados o agrupados correctamente, el cual fue de 91.3%. Pero aunque este indicador aumentó considerablemente con relación a los análisis anteriores, fue preciso observar los resultados de las pruebas estadísticas F y las correlaciones simples entre cada par de variables discriminantes. Se analizan las diferencias significativas de cada una de las variables originales, entre los dos grupos propuestos de estudiantes.

Los valores de la prueba o estadístico F nos muestra que las variables con mejor valoración son respectivamente **Perdidas** (12.172), **Padrespend** (5.348) y **Cartracasa** (4.028). Las correlaciones de estas variables son **Perdidas y Padrespend (0.044); Padrespend y Cartracasa (-0,075); Perdidas y Cartracasa (-0.024),** estos valores indican correlaciones muy débiles para estas variables pero se tomaran en cuenta, por la importancia en el proceso académico.

Las tres variables discriminantes que se tomaran en cuenta son, **Perdidas**, **Padrespend** y **Cartracasa;** ahora, con estas variables se procede a realizar el análisis estadístico discriminante definitivo; o sea que el análisis partió de 14 variables y se redujo a 3 variables buscando resultados más eficientes.

3.1.4. Análisis para 2 grupos y 3 variables predictoras.

Para este caso y al igual que los casos anteriores la variable grupo es Nota promedio (Nivel de competencia Bajo y Aprobado) la cual permite clasificar los datos, se determina trabajar con dos grupos de estudiantes, además se toma 3 variables predictoras para este análisis:

G1 (rendimiento Bajo nota promedio $1{,}0 < X \leq 5{,}9$): notas desde 1,0 hasta 5,9.

G2 (rendimiento Aprueba nota promedio $6{,}0 \leq X \leq 10{,}0$): notas desde 6,0 hasta 10,0.

G1 y G2 indica los grupos de pertenencia.

Para este análisis discriminante final se toman en cuenta sólo tres variables, las cuales son: **Perdidas** (Número de asignaturas perdidas), **Padrespend** (Padres pendiente del desarrollo de sus actividades académicas) y **Cartracasa** (Le ponen excesiva carga de trabajo en su casa).

Los resultados son los siguientes, y a continuación su descripción:

Tabla 37. Media y Desviación estándar de la variable en cada grupo.

Estadísticas de grupo					
Nivel1		Media	Desviación estándar	N válido (por lista)	
				No ponderados	Ponderados
Bajo	Perdidas	1,000	0,000	4	4
	Padrespend	2,250	1,258	4	4
	Cartracasa	3,250	2,062	4	4
Aprueba	Perdidas	2,310	0,750	180	180
	Padrespend	3,710	1,245	180	180
	Cartracasa	1,980	1,230	180	180
Total	Perdidas	2,280	0,766	184	184
	Padrespend	3,670	1,260	184	184
	Cartracasa	2,010	1,259	184	184

1. **Promedios para cada grupo**

En la Tabla 37, puede observarse que el grupo Aprueba tiene mayores valores de medias en sus variables, como se ve en dos de las variables: **Perdidas (2.310)** y **Padrespend (3.710)** en donde encontramos los valores más altos; este es un resultado que puede darse por la relación existente entre el número de asignaturas perdidas, el hecho de que los Padres estén pendiente del desarrollo de las actividades académicas de los estudiantes y rendimiento académico en general.

2. **Desviación estándar para cada grupo**

Por otro lado puede observarse que el grupo Aprueba en el Tabla 37, también tiene valores más bajos en la desviación estándar de sus variables, como se indica, en dos de las variables: **Cartracasa (1.230)** y **Padrespend (1.245)**; este es un resultado que puede darse por el hecho de que hay mayor cantidad de datos y que los datos son más parecidos o sea de menor diferencia.

Si hacemos un análisis de las variables en este caso encontramos que: En las variables **Perdidas** y **Padrespend**, las medias son más altas en el grupo Aprueba, mientras que la variable **Cartracasa** la media es más alta en el grupo Bajo, debido a que es más frecuente encontrar y asociar estudiantes de bajo rendimiento con el hecho de que realizan otras actividades laborales en su casa, restándole tiempo al estudio.

En cuanto a la variable **Perdidas** vemos que su desviación estándar es más alta en el grupo Aprueba, debido a que puede ser menos frecuente encontrar estudiantes perdiendo áreas en este grupo, por lo que hace creer que los pocos estudiantes que pierden asignaturas o áreas se convierten en sesgos de la muestra y por ende elevan la desviación estándar.

Tabla 38. Matrices dentro de grupos combinados

La matriz de covarianzas tiene 182 grados de libertad.

Matrices dentro de grupos combinados[a]				
		Perdidas	**Padrespend**	**Cartracasa**
Covarianza	Perdidas	0,553		
	Padrespend	0,041	1,550	
	Cartracasa	-0,022	-0,116	1,559
Correlación	Perdidas	1,000		
	Padrespend	0,044	1,000	
	Cartracasa	-0,024	-0,075	1,000

En el proceso de análisis se han eliminado variables con un alto grado de correlación porque se ha llegado a la conclusión de que no son tan eficientes o importantes en el proceso discriminante para lo cual se requieren o también porque los cálculos así lo han considerado; indicando así que la presencia de éstas variables pueden generar ciertas limitaciones en el análisis. De acuerdo a esto se ha decidido dejar las variables **Perdidas, Padrespend** y **Cartracasa** que tienen correlaciones poco significantes, pero que presentan características relevantes para asociarlas con el estudio del rendimiento académico.

La covarianza, Tabla 39, está mostrando como se relacionan los datos de acuerdo a su tamaño o sea que la dependencia directa es positiva cuando en una variable los datos son grandes y en la otra correlacionada también; de otro lado la dependencia directa es negativa si en una

variable los datos son grandes y en la otra correlacionada no lo son. O sea que las variables **Perdidas y Padrespend** tienen dependencia directa positiva, mientras que la variable **Cartracasa** correlacionada con las otras dos variables (Perdidas y Padrespend) tiene dependencia directa negativa.

Tabla 39. Matrices de covarianzas de grupos.

La matriz de covarianzas total tiene 183 grados de libertad.

Matrices de covarianzas[a]				
Nivel1		Perdidas	Padrespend	Cartracasa
Bajo	Perdidas	0,000		
	Padrespend	0,000	1,583	
	Cartracasa	0,000	1,917	4,250
Aprueba	Perdidas	0,562		
	Padrespend	0,042	1,550	
	Cartracasa	-0,023	-0,150	1,514
Total	Perdidas	0,586		
	Padrespend	0,082	1,587	
	Cartracasa	-0,058	-0,155	1,585

Ya en el caso anterior se ha explicado en que consiste la covarianza y como funciona en la correlación de dos variables cuando es positiva y negativa respectivamente, solo resta decir que los valores de cero (0) en la Tabla 39, indican que no hay covarianza o que no existe una dependencia directa entre las variables que se correlacionan o sea que no existencia de una relación lineal entre las dos variables correlacionadas.

Tabla 40. Prueba de igualdad de medias de grupos, Lambda de Wilks y Razón F Univariante.

Prueba de igualdad de medias de grupos					
	Lambda de Wilks	**F**	**gl1**	**gl2**	**Sig.**
Perdidas	0,937	12,172	1	182	0,001
Padrespend	0,971	5,348	1	182	0,022
Cartracasa	0,978	4,028	1	182	0,046

Mientras menor sea la Lambda de Wilks, mayor es el valor correspondiente de F y más altas son las posibilidades de que las medidas de los grupos sean significativamente diferentes. En la Tabla 40, puede observarse que la variable **Perdidas** tiene el menor valor de Lambda de Wilks y por supuesto, el mayor valor F demostrando con ello ser la variable que produce las diferencias más significativas entre los dos grupos que se estudian; además, es de considerar que el valor Lambda es muy alto para las tres variables lo que indica que hay un alto grado de solapamiento entre los datos de las variables que se correlacionan o sea que puede haber parecido entre los grupos. El valor de significancia es bueno, menor que 0.05 ($p \leq 0.05$), lo cual permite rechazar la hipótesis de igualdad de medias entre grupos y está dentro de lo permitido para el análisis.

3.1.4.1. Determinación del número de funciones discriminantes.

La aplicación $min\ (g-1,\ p)$ indica el mínimo entre p y $g-1$ lo cual muestra el número de funciones discriminantes debe ser tomadas en cuenta en este proceso de análisis, donde se utilizan dos grupos o sea que $g = 2$ y se toman en cuenta tres variables predictoras o sea que $p = 3$. Se expresa de la siguiente forma:

$min(g-1,\ p)$ Para $g = 2$, $P = 3$

$min(2-1,\ 3) = 1$ Se obtiene una sola función discriminante.

Tabla 41. Resumen de funciones discriminantes canónicas.
Se utiliza las primera 1 función discriminante canónica en el análisis.

Autovalores				
Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	0,110	100	100	0,314

La correlación canónica que muestra en la Tabla 41, es un valor de 0.314 lo cual es un dato relativamente bajo lo que indica que las variables discriminantes no permiten diferenciar lo suficientemente bien entre los grupos. Por otro lado el Autovalor que se han obtenido, reflejan un dato muy cercano a cero (0) por lo que se reafirma que las variables discriminantes utilizadas no permiten distinguir demasiado bien entre los 2 grupos; además, para este caso se muestra un porcentaje de varianza de discriminación equivalente a 100%, lo cual refleja que la función puede explicar el 100% de las diferencias existentes entre los datos de los grupos.

Tabla 42. Lambda de Wilks.

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	0,901	18,788	3	0,000

En el Tabla 42, el valor de lambda es alto (0.901) para la función, lo cual significa que existe buen solapamiento entre los grupos y el valor transformado de lambda (Chi-cuadrado = 18.788) tiene asociado con 3 grados de libertad, un nivel crítico (Sig.) de 0.000, por lo que podemos rechazar la hipótesis nula de que los grupos comparados tienen promedios iguales en las variables discriminantes en cuestión; aunque, estos valores indican que los grupos pueden parecerse un poco en las variables discriminantes en cuestión. Por tal motivo, para probar la significancia global de la función discriminante se planteó la hipótesis nula si las medias poblacionales son iguales, entonces en la tabla, Se verificó que el valor Chi-cuadrado (χ^2 =18.788) tiene un p valor menor de 0,05 (p=0,000), por consiguiente, se rechazó la hipótesis nula y se afirmó la significancia de la función discriminante.

Tabla 43. Centroides de grupo.

Funciones en centroides de grupo	
Nivel1	Función
	1
Bajo	-2,210
Aprueba	0,049

El resultado de la prueba Lambda de Wilks, o de su homólogo Chi-cuadrado (χ^2), indica la existencia de diferencia significativa entre los dos centroides de grupos. Esta Tabla 43 contiene la ubicación de los centroides en la función discriminante; esta, es de gran utilidad para interpretar la función. Podemos ver que el grupo Aprueba se encuentra localizado, en promedio, en las puntuaciones positivas de la función, mientras que el grupo Bajo se encuentran ubicados en las puntuaciones negativas de la función; por tal motivo, si desconocemos la procedencia de un individuo (estudiante) pero tenemos su información sobre las variables **Perdidas,**

Padrespend y **Cartracasa**, podemos calcular su puntuación discriminante y, a partir de ella, asignarlo al grupo de cuyo centroide se encuentre más próximo[18].

Tabla 44. Coeficientes de la función discriminante canónica (no tipificados)

	Función
	1
Perdidas	1,010
Padrespend	0,365
Cartracasa	-0,318
(Constante)	-3,008

3.1.4.2. Función canónica discriminante (no tipificada)

Coeficientes de la función canónica discriminante (no tipificada), Tabla 44, son los coeficientes utilizados por el programa (SPSS) para calcular las puntuaciones discriminantes y la ubicación de los centroides de los grupos. Por ejemplo, puede comprobarse que a partir de las medias de cada grupo en las variables discriminantes y este conjunto de coeficientes se obtienen los centroides en la función discriminante.

$$\bar{d}_1 = b_0 + b_1\bar{x}_1{}^{(1)} + b_2\bar{x}_2{}^{(1)} + b_3\bar{x}_3{}^{(1)}$$

$$\bar{d}_2 = b_0 + b_1\bar{x}_1{}^{(2)} + b_2\bar{x}_2{}^{(2)} + b_3\bar{x}_3{}^{(2)}$$

Donde: $\bar{d}_1$ y $\bar{d}_2$ Son los centroides de la función discriminante

b_0, b_1, b_2, b_3 Son los coeficientes de la función discriminante (no tipificados)

$\bar{x}_1{}^{(1)}, \bar{x}_2{}^{(1)}, \bar{x}_3{}^{(1)}$ Media de cada una de las variables en el grupo 1

$\bar{x}_1{}^{(2)}, \bar{x}_2{}^{(2)}, \bar{x}_3{}^{(2)}$ Media de cada una de las variables en el grupo 2

$$\bar{d}_1 = b_0 + b_1\bar{x}_1{}^{(1)} + b_2\bar{x}_2{}^{(1)} + b_3\bar{x}_3{}^{(1)}$$

$\bar{d}_1$ = -3,008 + 1,010 x 1,000 + 0,365 x 2,250 - 0,318 x 3,250 = -2,210

$$\bar{d}_2 = b_0 + b_1\bar{x}_1{}^{(2)} + b_2\bar{x}_2{}^{(2)} + b_3\bar{x}_3{}^{(2)}$$

$\bar{d}_2$ = -3,008 + 1,010 x 2,310 + 0,365 x 3,710 + -0,318 x 1,980 = 0,049

[18] Información complementada con Análisis discriminante: El procedimiento Discriminante. Capítulo 23

Verifíquese las medias de las variables de los dos grupos en la Tabla 37 y los valores de los centroides hallados por el programa SPSS en la Tabla 43 con el fin de identificar coincidencia en el cálculo.

Tabla 45. Coeficiente de la función canónica discriminante estandarizada y matriz de estructura.

	Coeficientes de función discriminante canónica estandarizados	Matriz de estructuras
Función	**1**	**1**
Perdidas	0,751	0,781
Padrespend	0,454	0,518
Cartracasa	-0,397	-0,449

En la Tabla 45, para la matriz de estructura se tiene en cuenta que las variables están ordenadas por el tamaño absoluto de la correlación dentro de la función, de acuerdo a esto vemos que en la Tabla 45, se da lugar a una función, la cual está asociada y correlacionada de la siguiente forma de mayor a menor: **Pérdidas** (0.781), **Padrespend** (0.518) y **Cartracasa** (-0.449); esto indica la relación bruta entre cada variable y la función discriminante; se representa que el valor negativo de la variables indica correlación negativa y por el contrario si es positiva; los valores altos se asignaron, en promedio, a los estudiantes con alto rendimiento. Por otro lado la variable que presentó una correlación negativa, indicando así que, en general, los valores altos de esta variable fueron atribuidos a los alumnos de más bajo rendimiento.

En la Tabla 45, se muestra los Coeficientes de función discriminante canónica estandarizados indica que para este caso la variable **Perdidas** (0.751) correlaciona en un nivel alto, la variable **Padrespend** (0.454) correlaciona en un nivel moderado; además, teniendo en cuenta que estas dos variables representan mucha importantes en la función; por otro lado, la variable **Cartracasa** (-0.449) correlaciona en un nivel moderado negativo en la función discriminante, pero por su valor muestran ser una variables de menor importantes en la función, esto puede deberse a el tipo de variables con la cual se han relacionado. Se observa que la variable de mayor peso es **Perdidas** y que el valor positivo indica que hay una relación directa entre la función discriminante y la variable, esto indica que cuanto mayor es la variable más alto es el valor de D obtenido, caso contrario sucede con la variable negativa **Cartracasa.**

3.1.4.3. Función canónica estandarizada discriminante.

Conocidos los coeficientes de la función discriminante (Tabla 45), se construye la función D capaz de diferenciar lo más posible a los dos grupos proyectados en el presente análisis, que es una combinación lineal de ambas variables, ya que se representa a partir de los coeficientes de la tabla de los Coeficientes de función discriminante canónica estandarizados generados por el cálculo del programa (SPSS). La función discriminante obtenida en el presente análisis es la siguiente:

$$D = 0{,}751^{*}\ \text{Perdidas} + 0{,}454^{*}\ \text{Padrespend} - 0{,}397^{*}\ \text{Cartracasa}$$

Los valores decimales de la fórmula muestran las ponderaciones de las variables independientes que consiguen hacer que los sujetos de uno de los grupos obtengan puntuaciones máximas en D, y los sujetos del otro grupo puntuaciones mínimas. Además, **Perdidas**, **Padrespend** y **Cartracasa** corresponde a los valores de la variable discriminante que me pueden aproximar el dato a uno de los grupos, el cual puede ser el más cercano al centroide más próximo.

Tabla 46. Coeficientes de función de clasificación.

Funciones discriminantes lineales de Fisher

Coeficientes de función de clasificación		
	Nivel1	
	Bajo	Aprueba
Perdidas	1,783	4,065
Padrespend	1,571	2,395
Cartracasa	2,228	1,509
(Constante)	-6,971	-11,325

3.1.4.4. Funciones discriminantes lineales de Fisher

$$F_1 = -6{,}971 + 1{,}783^{*}\ \text{Perdidas} + 1{,}571^{*}\ \text{Padrespend} + 2{,}228^{*}\ \text{Cartracasa}$$

$$F_2 = -11{,}325 + 4{,}065^{*}\ \text{Perdidas} + 2{,}395^{*}\ \text{Padrespend} - 1{,}509^{*}\ \text{Cartracasa}$$

En este caso las fórmulas muestran las funciones discriminantes lineales de Fisher, donde permite reemplazar los valores de las variables de un individuo en especial, generando resultados que nos llevan a clasificar el individuo en el grupo 1 o grupo 2 dependiendo cuan cerca este el resultado del centroide más cercano.

Tabla 47. Resultados de clasificación.

86,4% de casos agrupados originales clasificados correctamente.

		Nivel1	Pertenencia a grupos pronosticada		Total	
			Bajo	Aprueba		
Original	Recuento	Bajo	4	0	4	184
		Aprueba	25	155	180	
	%	Bajo	100%	0,0%	100%	
		Aprueba	13,9%	86,1%	100%	

La Tabla 47 muestra que el porcentaje de correcta clasificación fue del 86.4% lo cual es un alto porcentaje, pero disminuyo 4.9% con relación al análisis anterior, donde se utilizaron 2 grupos y 14 variables predictoras; es importante tener en cuenta que el análisis se hace más eficiente al reducir las variables; en este punto radica la importancia de este último proceso, ya que se utilizaron un menor número de variable predictoras conservándose el número de grupos (2 grupo y 3 variables).

El grupo, Bajo, clasifica correctamente el 100% de sus estudiantes y el grupo, Aprueba, el 86.1%, pero el porcentaje no se ve más elevado dado los pocos estudiantes que hay en el grupo Bajo; además, se identifica que el porcentaje de mala clasificación es del grupo Aprueba.

Si desea calcular la probabilidad a posteriori $(P_r(g/D))$ se utiliza la expresión:

$$\left(P_r\left(\frac{g}{D}\right)\right) = \frac{e^{F}g}{e^{F}1+e^{F}2}$$

Esta probabilidad a partir de la puntuación discriminante (D) clasifica a los estudiantes en el grupo 1 o 2, el individuo será clasificado en el grupo que tenga la mayor probabilidad a posteriori.

Ejemplo: Un estudiante con las siguientes variables.

Perdidas = 2 (regular), Padrespend = 3 (algunas veces), Cartracasa = 2 (muy pocas veces)

¿Dónde se clasificará?

$$F_1 = -6{,}971 + 1{,}783^* \text{ Perdidas} + 1{,}571^* \text{ Padrespend} + 2{,}228^* \text{ Cartracasa},$$

$$F_2 = -11{,}325 + 4{,}065^* \text{ Perdidas} + 2{,}395^* \text{ Padrespend} - 1{,}509^* \text{ Cartracasa},$$

$$F_1 = -6{,}971 + 1{,}783^* 2 + 1{,}571^* 3 + 2{,}228^* 2 = 4{,}564,$$

$$F_2 = -11{,}325 + 4{,}065^* 2 + 2{,}395^* 3 - 1{,}509^* 2 = 0{,}972,$$

$$\left(P_r(g = 1/D)\right) = \frac{e^{4{,}564}}{e^{4{,}564} + e^{0{,}972}} = 0{,}9732,$$

$$\left(P_r(g = 2/D)\right) = \frac{e^{0{,}972}}{e^{4{,}564} + e^{0{,}972}} = 0{,}0268.$$

Esto nos indica que el nuevo estudiante escogido tiene mayor probabilidad de ubicarse en el grupo 1 (Bajo).

3.2. Método kernel, aplicación y resultados

3.2.1. Clasificación de individuos.

Para construir un estimador no lineal, debemos transformar los datos de entrada no linealmente. La transformación no lineal implica una correspondencia hacia un espacio de mayor dimensión, posiblemente infinita. Existe un producto escalar en ese espacio que pueda ser expresado como función de los datos de entrada x . El producto escalar explícito es en el método Kernel

• Existe una expresión del producto escalar en función del espacio de entrada.

• No necesitamos la expresión de las componentes del vector en el espacio de características. Esta transformación incrementa la posibilidad de que haya separabilidad lineal. Este es un ejemplo de correspondencia en un espacio de Hilbert. En una dimensión, no es posible clasificar los datos linealmente. Todo esto mediante la premisa que cualquier modelo lineal puede convertirse en un modelo no lineal aplicando el truco del kernel al modelo reemplazando sus características (predictores) por una función.

Recordemos que no se necesitan los vectores, así que no debemos preocuparnos por la dimensión del espacio en cuanto a costo computacional. Lo que necesitamos es conocer el kernel.

En esta parte del estudio se hace la aplicación del kernel de tal manera que permita hacer el comparativo e identificación de la eficiencia en la clasificación de individuos en una muestra; para tal caso se divide la muestra en 4, 3 y 2 grupo, teniendo en cuenta la variable grupo llamada **Nota promedio**. El hecho de dividir la muestra en 4, 3 y 2 grupo, lo que busca es generar condiciones similares a las aplicadas con el método de Análisis Discriminante, con el fin de que el comparativo se haga más claro y eficaz.

En el método kernel, como ya se sabe por conceptos anteriores tiene varias fórmulas que se pueden aplicar dependiendo para lo cual se requieran o la comodidad para los cálculos.

Los kernel son llamados medidas de similaridad y buscando la mejor aplicación en este caso, se hace uso durante todo el proceso de la fórmula del kernel Gaussiano la cual se muestra a continuación:

$$K(x_i, \bar{x}_j) = \exp\left(\frac{-\| x_i - \bar{x}_j \|^2}{2\sigma^2}\right).$$

Donde K representa el kernel, x_i es un dato determinado de la serie, $\bar{x}_j$ es la media de los datos del grupo o de la muestra, σ^2 es la varianza, el cual es un dato estimado desde 0.01 hasta 1.0 buscando la mejor clasificación.

En la Tabla 48, se indica la subdivisión de la muestra en grupos, y para cada caso los intervalos que dividen los grupos de acuerdo a la variable grupo, el número de individuos por grupo y total, el promedio por grupo y total; además, la varianza por grupo y total. Estos datos se utilizan más adelante con el kernel.

Tabla 48. Datos descriptivos para el kernel.

Grupo	Intervalo		Conteo	Prom. Grupo	Varianza
Bajo	1,0	5,9	4	5,7750	0,0225
Básico	6,0	7,9	156	6,8699	0,2323
Alto	8,0	8,9	20	8,2700	0,0443
Superior	9,0	10,0	4	9,1250	0,0358
Total			184	7,0473	0,5223
Grupo	Intervalo		Conteo	Prom. Grupo	Varianza
Bajo	1,0	5,9	4	5,7750	0,0225
Medio	6,0	8,0	158	6,8842	0,2454
Alto	8,1	10,0	22	8,4125	0,1472
Total			184	7,0473	0,5223
Grupo	Intervalo		Conteo	Prom. Grupo	Varianza
Bajo	1,0	5,9	4	5,775	0,0225
Aprueba	6,0	10,0	180	7,0756	0,4966
Total			184	7,0473	0,5223

Como ya lo habíamos dicho, los kernel son medidas de similaridad y en especial el kernel Gaussiano que se mueven en un intervalo de 0 a 1, donde 1 (uno) es el nivel más alto de similaridad y 0 (cero) es el menor nivel de similaridad, o sea, mayor diferencia, esto se logró notar en la práctica con los datos de la muestra.

Para la aplicación del kernel se hace uso de la media de los grupos o de la media de la muestra total, los valores de los datos estimados de la varianza en los grupos y varianzas de

ensayo desde 0.01 hasta 1.00 buscando identificar el valor de σ^2 que determine la mejor clasificación de los datos.

A continuación se muestra en la Tabla 49, para 4 grupos (Bajo, Básico, Alto y Superior), los datos obtenidos por la función kernel y clasificados por el valor de la varianza estimada y de ensayo similar al grupo, esto se determina al clasificar los valores más altos desde 0.6 hasta 1.0, ya que en este punto empieza a mostrar una significancia de similaridad que puede aportar la base para clasificar el dato o individuo al grupo determinado. En la parte superior de la Tabla 49, y en correlación (Grupo vs Varianza), vemos el número de individuos bien clasificados para cada grupo y varianza determinada como se indica a continuación según la numeración que aparece en la parte baja:

1. **De cada grupo:** Aquí se utiliza las Varianzas y las medias de cada grupo.
2. **De la muestra:** Aquí se utiliza la Varianza muestral y la media muestral.
3. **VM. De la muestra y Gru.:** Aquí se utiliza las Varianza muestral y las medias de cada grupo.
4. Varianzas de ensayo 0.01 y la media de cada grupo.
5. Varianzas de ensayo 0.02 y la media de cada grupo.
6. Varianzas de ensayo 0.03 y la media de cada grupo.
7. Varianzas de ensayo 0.04 y la media de cada grupo.
8. Varianzas de ensayo 0.05 y la media de cada grupo.
9. Varianzas de ensayo 0.7 y la media de cada grupo.
10. Varianzas de ensayo 1.0 y la media de cada grupo.

En la parte inferior de la Tabla 49, y sobre cada grupo se muestra el porcentaje de individuos bien clasificados, además, en el lugar indicado como Total vemos el número de individuos y el porcentaje total de individuos bien clasificados, de lo cual podemos decir que la mayor eficiencia de clasificación la vemos para una varianza de 0.5 (VM) a 0.7, con 93.5% respectivamente en cada caso utilizando la media de cada grupo. Se hace la aclaración que no se asume la Varianza 1.0 como mejor clasificación ya que está arrojando un valor que es muy improbable, indicando el 100% de buena clasificación para la muestra.

Tabla 49. Resultados de clasificación kernel, 4 grupos.

	Varianza									
Grupo	De cada Grupo	De la muestra	VM. De la muestra y Gru.	0,01	0,02	0,03	0,04	0,05	0,7	1,0
Bajo	3	0	4	1	3	4	4	4	4	4
Básico	105	122	144	22	30	39	54	54	144	156
Alto	14	0	20	7	9	14	14	16	20	20
Superior	3	0	4	1	3	3	3	3	4	4
Total	125	122	172	31	45	60	75	77	172	184
Bajo	75,0%	0,0%	100,0%	25,0%	75,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Básico	67,3%	66,3%	92,3%	14,1%	19,2%	25,0%	34,6%	34,6%	92,3%	100,0%
Alto	70,0%	0,0%	100,0%	35,0%	45,0%	70,0%	70,0%	80,0%	100,0%	100,0%
Superior	75,0%	0,0%	100,0%	25,0%	75,0%	75,0%	75,0%	75,0%	100,0%	100,0%
Total	67,9%	66,3%	93,5%	16,8%	24,5%	32,6%	40,8%	41,8%	93,5%	100,0%
	1	2	3	4	5	6	7	8	9	10

Para la Tabla 50, encontramos 3 grupos (Bajo, Básico y Alto) con las mismas varianzas que en la tabla anterior y las medias aplicadas de la misma forma. En este caso volvemos a encontrar la mayor eficiencia de clasificación en la varianza de 0.5 a 0.7, con 91.8% respectivamente en cada caso utilizando la media de cada grupo. Aunque para la varianza 1.0 el valor disminuyo con 3 grupos, se debe tener cierto cuidado o cautela, pues el valor aun es alto (97.8%) y puede dar lugar a una mala clasificación de algunos individuos para la muestra (datos hallados con Excel).

Tabla 50. Resultados de clasificación kernel, 3 grupos.

	Varianza									
Grupo	De cada Grupo	De la muestra	VM. De la muestra y Gru.	0,01	0,02	0,03	0,04	0,05	0,7	1,0
Bajo	3	0	4	1	3	4	4	4	4	4
Básico	105	122	144	22	30	30	39	54	144	154
Alto	18	0	21	4	4	8	8	8	21	22
Total	126	122	169	27	37	42	51	66	169	180
Bajo	75,0%	0,0%	100,0%	25,0%	75,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Básico	66,5%	66,3%	91,1%	13,9%	19,0%	19,0%	24,7%	34,2%	91,1%	97,5%
Alto	81,8%	0,0%	95,5%	18,2%	18,2%	36,4%	36,4%	36,4%	95,5%	100,0%
Total	68,5%	66,3%	91,8%	14,7%	20,1%	22,8%	27,7%	35,9%	91,8%	97,8%

Para la Tabla 51, utilizamos 2 grupos (Bajo y Aprobado) con las mismas varianzas que en la tabla anterior y las medias aplicadas de la misma forma. En este caso encontramos la mayor eficiencia de clasificación en la varianza de 0.7 a 1.0, con 81.5% y 87.5% respectivamente en cada caso utilizando la media de cada grupo.

Tabla 51. Resultados de clasificación kernel, 2 grupos.

	Varianza									
Grupo	De cada Grupo	De la muestra	VM. De la muestra y Gru.	0,01	0,02	0,03	0,04	0,05	0,7	1,0
Bajo	3	0	4	1	3	4	4	4	4	4
Aprueba	122	122	131	23	30	30	38	45	146	157
Total	125	122	135	24	33	34	42	49	150	161
Bajo	75,0%	0,0%	100,0%	25,0%	75,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Aprueba	67,8%	67,8%	72,8%	12,8%	16,7%	16,7%	21,1%	25,0%	81,1%	87,2%
Total	67,9%	66,3%	73,4%	13,0%	17,9%	18,5%	22,8%	26,6%	81,5%	87,5%

En la Tabla 52, clasifica la muestra en dos grupos (Bajo y Aprobado) e indica el tamaño de cada grupo, el promedio de los grupos y de la muestra; además, pondera el valor del centroide en cada grupo con la aplicación del kernel Gaussiano y también se indica el punto de corte discriminante, esto con el fin de poder clasificar los individuos por su cercanía al centroide determinado para cada grupo, dependiendo la probabilidad por tamaño de grupo.

Punto de corte discriminante para muestras de igual tamaño $C = \frac{\bar{x}_1 + \bar{x}_2}{2}$.

Punto de corte discriminante para muestras de diferente tamaño $C = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.

Dónde: C Es el punto de corte discriminante

$\bar{x}_1$ Media aritmética del grupo 1

$\bar{x}_2$ Media aritmética del grupo 2

n_1 Tamaño del grupo 1

n_2 Tamaño del grupo 2

Tabla 52. Datos descriptivos con ponderación del centroide, 2 grupos

Punto de corte Discriminante	Centroides Promedio Kernel	Grupo	Intervalo		Conteo	Promedio grupo
C = 0,15392	0,4716929	Bajo	1,0	5,9	4	5,7750
	0,1468628	Aprueba	6,0	10,0	180	7,0756
Centroide	0,1539243	Total			184	7,0473

En la Tabla 53, se muestra el ejemplo de clasificación de individuos de la muestra utilizando el kernel de acuerdo a la cercanía al centroide del grupo, con una varianza de 0.01 y un 22.8% de correcta clasificación. Para esto se toma en cuenta la probabilidad para cada grupo (datos hallados con Excel):

Probabilidad para el grupo 1, n_1 $\quad P(n_1) = \frac{n_1}{n_1 + n_2}$.

Probabilidad para el grupo 2, n_2 $\quad P(n_2) = \frac{n_2}{n_1 + n_2}$.

Tabla 53. Ejemplo de clasificación de datos con ponderación del centroide.

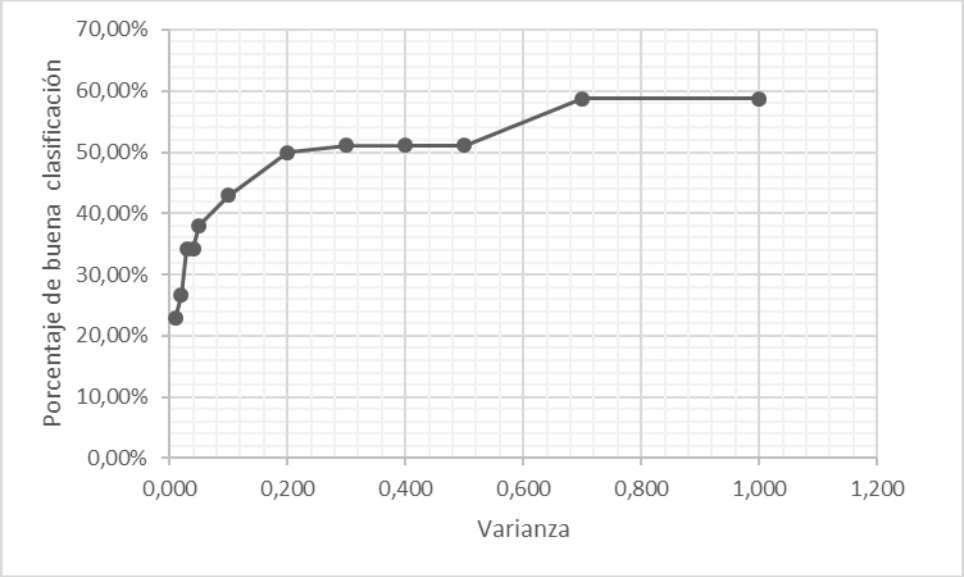
Ejemplo para Varianza 0,01		
Grupo	Bajo	Aprueba
Bajo	4	0
Aprueba	142	38
Total	146	38
Bajo	100,0%	0,0%
Aprueba	78,9%	21,1%
Total Bien clasificado	22,8%	

En la Tabla 54 se muestran los porcentajes de buena clasificación por el método cercanía al centroide más próximo, con varianzas desde 0.01 hasta 1.0, aquí vemos reflejada una proporcionalidad casi que directa entre la varianza y el porcentaje; además, la mayor clasificación está dada en la varianza 0.70 la cual es de 58.7%, valor de buena clasificación que se conserva en datos modelados hasta más de 3 puntos de varianza, pero su incremento no es muy significativo, casi que conservando una línea que no pasa del 61% (ver figura 4).

Tabla 54. Clasificados correctamente de acuerdo a las Varianzas y con ponderación del centroide. Diseño propio

Clasificados correctamente de acuerdo a las varianzas			
Varianza	Porcentaje de buena clasificación	Varianza	Porcentaje de buena clasificación
0,01	22,83%	0,20	50,00%
0,02	26,63%	0,30	51,09%
0,03	34,24%	0,40	51,09%
0,04	34,24%	0,50	51,09%
0,05	38,04%	0,70	58,70%
0,10	42,93%	1,00	58,70%

Figura 4: Gráfica de Clasificados correctamente de acuerdo a las varianzas y ponderación del centroide. Diseño propio.



Si comparamos los dos procesos que se realizaron para los dos grupos, (el primero utilizando solo la similaridad que genera el kernel y el segundo, teniendo en cuenta la probabilidad que se crea en cada grupo por su diferencia de tamaño) podemos encontrar que tiene más eficiencia de clasificación el primer proceso, ya que en este la buena clasificación alcanza más del 80% y en el segundo no pasa del 61%.

3.2.2. Discriminación de las variables utilizando kernel y comparativo con las varianzas

Usando el criterio de similaridad dado por los kernel (donde 1 es muy similar y 0 no hay similaridad o la similaridad es nula), se usa este criterio para compararlo con el estadístico Lambda de Wilks, la cual expresa la proporción de variabilidad total de las variables en los grupos, o sea que cuando los grupos se encuentran superpuesto en el espacio multidimensional los valores del numerador y denominador serán aproximadamente iguales o sea 1 y en la medida que los grupos se van separando aumenta la variabilidad intragrupos y al comparar, la variabilidad total disminuye (haciéndose menor el valor del cociente)[19].

Por lo anterior, podemos decir que al hallar la variabilidad total de una variable con la función kernel (función kernel Gaussiano), el valor obtenido más cercano a 0 (cero) demuestra mayor variabilidad y por tanto menor similaridad, lo que refleja mayor capacidad de esa variable par la discriminación de datos (mayor poder discriminante).

En las Tabla 55 a 60, se realiza el comparativo para la discriminación de variables usando por la función kernel, para esto se usa los valores de la desviación estándar y se analiza cómo puede variar estos datos usando las medias en cada variable. Es de aclarar que este proceso se realiza para 4, 3 y 2 grupos; para cada caso se hace una modelación con la varianza como se muestra en la Tabla 55, en cada división de grupos de la muestra (se utiliza el programa Excel). Esto permite que en la medida como se le van dando valores a la varianza se van haciendo cambios en el resultado final.

La varianza utilizada en la modelación para este estudio está dada en un intervalo entre 0,01 y 1,0 esto permite ver cuáles son las variables de mayor discriminación (más cercanas a cero) y cómo cambia el dato con la varianza, mostrando cuan sensible puede ser este estadístico.

[19] Análisis discriminante. Capítulo 23

Para 4 grupos encontramos los siguientes datos:

Tabla 55. Ponderación para identificar el poder discriminante de las variables modelando el kernel. 4 grupos

Desviación estándar						Varianza
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**	0,01
Perdidas	0,000	0,751	0,000	0,000	0,766	
Tiempodedif	1,258	0,931	1,196	0,500	0,975	
Entregatareas	1,500	0,829	1,214	0,577	0,911	
Padrespend	1,258	1,249	1,294	0,957	1,260	
Convives	1,000	0,679	0,733	0,577	0,685	
Ambientefam	0,957	1,168	1,268	0,816	1,164	
Educacionrec	1,000	1,075	1,281	0,577	1,084	
Tegustacol	1,893	1,124	1,164	0,500	1,132	
Haspenfu	0,000	1,076	1,302	0,000	1,083	
Vicbullyng	0,500	1,298	1,518	0,000	1,300	
Nivelespa	0,500	0,598	0,718	0,500	0,605	
Cartracasa	2,062	1,229	1,373	0,500	1,259	
Respetocodoc	1,414	1,246	0,923	0,816	1,216	
Pactoconv	2,062	1,106	1,099	0,957	1,123	

Matriz de similaridad D.E.					
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**
Perdidas	0,000	0,989	0,000	0,000	**0,838**
Tiempodedif	0,018	0,908	0,087	0,000	**0,779**
Entregatareas	0,000	0,714	0,010	0,004	**0,607**
Padrespend	1,000	0,994	0,944	0,010	0,967
Convives	0,007	0,998	0,891	0,558	0,955
Ambientefam	0,117	0,999	0,582	0,002	0,913
Educacionrec	0,703	0,996	0,144	0,000	0,875
Tegustacol	0,000	0,997	0,950	0,000	0,948
Haspenfu	0,000	0,998	0,091	0,000	**0,856**
Vicbullyng	0,000	1,000	0,093	0,000	**0,858**
Nivelespa	0,576	0,998	0,528	0,576	0,928
Cartracasa	0,000	0,956	0,522	0,000	**0,867**
Respetocodoc	0,141	0,956	0,014	0,000	**0,815**
Pactoconv	0,000	0,986	0,972	0,252	0,947

Medias					
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**
Perdidas	1,000	2,210	3,000	3,000	2,280
Tiempodedif	2,750	2,570	3,200	2,750	2,650
Entregatareas	3,250	3,440	4,000	4,500	3,520
Padrespend	2,250	3,680	3,900	3,750	3,670
Convives	3,500	3,400	3,300	3,500	3,390
Ambientefam	4,250	4,150	4,350	4,000	4,170
Educacionrec	4,500	4,180	4,200	4,500	4,200
Tegustacol	3,750	4,040	4,250	3,750	4,050
Haspenfu	5,000	4,400	4,300	5,000	4,420
Vicbullyng	1,250	1,750	1,900	1,000	1,740
Nivelespa	2,750	2,790	2,900	2,750	2,800
Cartracasa	3,250	2,000	1,900	1,750	2,010
Respetocodoc	4,000	3,760	4,300	4,000	3,830
Pactoconv	2,750	3,260	3,550	3,250	3,280

Matriz de similaridad de Medias					
Variables	**Bajo**	**Basico**	**Alto**	**Superior**	**Total**
Perdidas	0,000	0,783	0,000	0,000	**0,664**
Tiempodedif	0,607	0,726	0,000	0,607	**0,642**
Entregatareas	0,026	0,726	0,000	0,000	**0,616**
Padrespend	0,000	0,995	0,071	0,726	**0,867**
Convives	0,546	0,995	0,667	0,546	0,940
Ambientefam	0,726	0,980	0,198	0,236	**0,873**
Educacionrec	0,011	0,980	1,000	0,011	0,940
Tegustacol	0,011	0,995	0,135	0,011	**0,859**
Haspenfu	0,000	0,980	0,487	0,000	0,884
Vicbullyng	0,000	0,995	0,278	0,000	0,874
Nivelespa	0,882	0,995	0,607	0,882	0,948
Cartracasa	0,000	0,995	0,546	0,034	0,904
Respetocodoc	0,236	0,783	0,000	0,236	**0,674**
Pactoconv	0,000	0,980	0,026	0,956	**0,855**

De los datos de la Tabla 55, aparecen organizados en el Tabla 56 los valores de similitud (discriminante) del kernel para 4 grupos, la desviación estándar y la media, mostrando como varía para algunas estimaciones y ensayos de varianza, esto permite tener una idea de cuáles son las variables más discriminantes. En este caso vemos las 4 o 5 variables que tienen mayor discriminación para cada ejemplo de varianza y encuentra que las variables que muestran mayor

poder discriminante son **Perdidas**, **Entregatareas**, **Tiempodedif**, ya que están incluidas en mayor proporción en las diferentes estimaciones y en menor proporción **Padrespend**. Algunas variables como **Haspenfu** y **Vicbullyng** a pesar de estar incluidas en buena proporción en el análisis, en este caso no son consideradas ya que se cree que no tienen mucha incidencia en el actual estudio.

Tabla 56. Valores kernel de la media y la desviación para variables discriminantes según la varianza. 4 grupos.

varianza 0.01			varianza 0.05		
Variables	Desviación	Media	Variables	Desviación	Media
Entregatareas	0,607	0,616	Entregatareas	0,844	0,817
Tiempodedif	0,779	0,642	Perdidas	0,846	0,808
Respetocodoc	0,815	0,674	Respetocodoc	0,905	0,852
Perdidas	0,838	0,664	Tiempodedif	0,910	0,840

varianza 0.10			varianza 0.50		
Variables	Desviación	Media	Variables	Desviación	Media
Perdidas	0,855	0,837	Perdidas	0,932	0,926
Entregatareas	0,905	0,871	Haspenfu	0,965	0,986
Haspenfu	0,933	0,955	Vicbullyng	0,967	0,983
Vicbullyng	0,934	0,951	Entregatareas	0,976	0,957
Tiempodedif	0,946	0,886	Cartracasa	0,978	0,980

varianza 0.70			varianza 1.00		
Variables	Desviación	Media	Variables	Desviación	Media
Perdidas	0,948	0,942	Perdidas	0,961	0,956
Haspenfu	0,972	0,989	Haspenfu	0,978	0,992
Vicbullyng	0,973	0,988	Vicbullyng	0,979	0,991
Entregatareas	0,983	0,968	Cartracasa	0,987	0,987
Cartracasa	0,983	0,983	Entregatareas	0,988	0,976

En la Tabla 57, se realiza el comparativo para la discriminación de variables utilizando 3 grupos y al igual que el anterior con la función kernel, se hace una modelación con la varianza. Para 3 grupos encontramos los siguientes datos:

Tabla 57. Ponderación para identificar el poder discriminante de las variables por el kernel. 3 grupos.

DESVIACIÓN ESTANDAR				
VARIABLES	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	0,000	0,752	0,000	0,766
Tiempodedif	1,258	0,934	1,066	0,975
Entregatareas	1,500	0,842	1,155	0,911
Padrespend	1,258	1,267	1,019	1,260
Convives	1,000	0,677	0,716	0,685
Ambientefam	0,957	1,167	1,211	1,164
Educacionrec	1,000	1,072	1,220	1,084
Tegustacol	1,893	1,122	1,109	1,132
Haspenfu	0,000	1,072	1,255	1,083
Vicbullyng	0,500	1,293	1,468	1,300
Nivelespa	0,500	0,595	0,710	0,605
Cartracasa	2,062	1,247	1,110	1,259
Respetocodoc	1,414	1,242	0,922	1,216
Pactoconv	2,062	1,108	1,057	1,123

Matriz de similaridad D.E.				
Variables	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	0,000	0,990	0,000	0,850
Tiempodedif	0,018	0,919	0,661	0,869
Entregatareas	0,000	0,788	0,051	0,683
Padrespend	1,000	0,998	0,055	0,885
Convives	0,007	0,997	0,953	0,970
Ambientefam	0,117	1,000	0,895	0,968
Educacionrec	0,703	0,993	0,397	0,915
Tegustacol	0,000	0,995	0,974	0,971
Haspenfu	0,000	0,994	0,228	0,881
Vicbullyng	0,000	0,998	0,244	0,886
Nivelespa	0,576	0,995	0,576	0,936
Cartracasa	0,000	0,993	0,330	0,892
Respetocodoc	0,141	0,967	0,013	0,835
Pactoconv	0,000	0,989	0,804	0,945

MEDIAS				
VARIABLES	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	1,000	2,220	3,000	2,280
Tiempodedif	2,750	2,560	3,230	2,650
Entregatareas	3,250	3,460	4,000	3,520
Padrespend	2,250	3,650	4,090	3,670
Convives	3,500	3,400	3,320	3,390
Ambientefam	4,250	4,150	4,320	4,170
Educacionrec	4,500	4,190	4,180	4,200
Tegustacol	3,750	4,050	4,090	4,050
Haspenfu	5,000	4,410	4,360	4,420
Vicbullyng	1,250	1,740	1,820	1,740
Nivelespa	2,750	2,800	2,860	2,800
Cartracasa	3,250	2,010	1,770	2,010
Respetocodoc	4,000	3,770	4,230	3,830
Pactoconv	2,750	3,270	3,450	3,280

Varianza
0,01

Matriz de similaridad de Medias				
Variables	**Bajo**	**Medio**	**Alto**	**Total**
Perdidas	0,000	0,835	0,000	0,717
Tiempodedif	0,607	0,667	0,000	0,586
Entregatareas	0,026	0,835	0,000	0,718
Padrespend	0,000	0,980	0,000	0,842
Convives	0,546	0,995	0,783	0,960
Ambientefam	0,726	0,980	0,325	0,896
Educacionrec	0,011	0,995	0,980	0,972
Tegustacol	0,011	1,000	0,923	0,969
Haspenfu	0,000	0,995	0,835	0,954
Vicbullyng	0,000	1,000	0,726	0,946
Nivelespa	0,882	1,000	0,835	0,978
Cartracasa	0,000	1,000	0,056	0,865
Respetocodoc	0,236	0,835	0,000	0,722
Pactoconv	0,000	0,995	0,236	0,883

De los datos anteriores, en la Tabla 57, aparecen organizados los valores de similitud (discriminante) del kernel para la desviación estándar y la media de 3 grupos, en la Tabla 58; vemos 4 o 5 variables que tienen mayor discriminación para cada varianza de ensayo utilizada;

se observa que las variables que muestran mayor poder discriminante para 3 grupos son **Perdidas**, **Entregatareas**; además, pueden ser aceptadas en menor ponderación **Padrespend**, **Tiempodedif**, ya que están incluidas en mayor proporción en las diferentes estimaciones. Otras variables a pesar de estar incluidas en buena proporción en el análisis, no son consideradas ya que se cree que no inciden en el actual estudio.

Tabla 58. Valores kernel de la media y la desviación para variables discriminantes según la varianza. 3 grupos.

varianza 0.01			varianza 0.05		
Variables	Desviación	Media	Variables	Desviación	Media
Entregatareas	0,683	0,718	Perdidas	0,857	0,829
Respetocodoc	0,835	0,722	Entregatareas	0,885	0,851
Perdidas	0,850	0,717	Respetocodoc	0,918	0,869
Tiempodedif	0,869	0,586	Padrespend	0,947	0,876

varianza 0.10			varianza 0.50		
Variables	Desviación	Media	Variables	Desviación	Media
Perdidas	0,865	0,852	Perdidas	0,937	0,931
Entregatareas	0,931	0,896	Haspenfu	0,981	0,993
Respetocodoc	0,951	0,916	Entregatareas	0,983	0,971
Haspenfu	0,961	0,980	Vicbullyng	0,986	0,995
Vicbullyng	0,963	0,981	Cartracasa	0,987	0,976

varianza 0.70			varianza 1.00		
Variables	Desviación	Media	Variables	Desviación	Media
Perdidas	0,952	0,946	Perdidas	0,964	0,959
Haspenfu	0,985	0,995	Haspenfu	0,989	0,996
Entregatareas	0,987	0,979	Entregatareas	0,991	0,985
Pactoconv	0,989	0,994	Pactoconv	0,992	0,995
Vicbullyng	0,990	0,996	Vicbullyng	0,992	0,997

En el Tabla 59 se realiza el comparativo para la discriminación de variables utilizando 2 grupos y al igual que el anterior con la función kernel, se hace la modelación con la varianza. Para 2 grupos encontramos los siguientes datos:

Tabla 59. Ponderación para identificar el poder discriminante de las variables por el kernel. 2 grupos.

DESVIACIÓN ESTANDAR			
VARIABLES	**Bajo**	**Aprobado**	**Total**
Perdidas	0,000	0,752	0,766
Tiempodedif	1,258	0,934	0,975
Entregatareas	1,500	0,842	0,911
Padrespend	1,258	1,267	1,260
Convives	1,000	0,677	0,685
Ambientefam	0,957	1,167	1,164
Educacionrec	1,000	1,072	1,084
Tegustacol	1,893	1,122	1,132
Haspenfu	0,000	1,072	1,083
Vicbullyng	0,500	1,293	1,300
Nivelespa	0,500	0,595	0,605
Cartracasa	2,062	1,247	1,259
Respetocodoc	1,414	1,242	1,216
Pactoconv	2,062	1,108	1,123

Matriz de similaridad D.E.			
Variables	**Bajo**	**Aprobado**	**Total**
Perdidas	0,0000	0,9902	0,969
Tiempodedif	0,0182	0,9194	0,900
Entregatareas	0,0000	0,7882	0,771
Padrespend	0,9998	0,9976	0,998
Convives	0,0070	0,9968	0,975
Ambientefam	0,1174	0,9996	0,980
Educacionrec	0,7027	0,9928	0,987
Tegustacol	0,0000	0,9950	0,973
Haspenfu	0,0000	0,9940	0,972
Vicbullyng	0,0000	0,9976	0,976
Nivelespa	0,5762	0,9950	0,986
Cartracasa	0,0000	0,9928	0,971
Respetocodoc	0,1408	0,9668	0,949
Pactoconv	0,0000	0,9888	0,967

	MEDIAS			Varianza
VARIABLES	**Bajo**	**Aprobado**	**Total**	0,01
Perdidas	1,000	2,310	2,280	
Tiempodedif	2,750	2,640	2,650	
Entregatareas	3,250	3,530	3,520	
Padrespend	2,250	3,710	3,670	
Convives	3,500	3,390	3,390	
Ambientefam	4,250	4,170	4,170	
Educacionrec	4,500	4,190	4,200	
Tegustacol	3,750	4,060	4,050	
Haspenfu	5,000	4,410	4,420	
Vicbullyng	1,250	1,750	1,740	
Nivelespa	2,750	2,810	2,800	
Cartracasa	3,250	1,980	2,010	
Respetocodoc	4,000	3,820	3,830	
Pactoconv	2,750	3,290	3,280	

Matriz de similaridad de Medias			
Variables	**Bajo**	**Aprobado**	**Total**
Perdidas	0,0000	0,9560	0,935
Tiempodedif	0,6065	0,9950	0,987
Entregatareas	0,0261	0,9950	0,974
Padrespend	0,0000	0,9231	0,903
Convives	0,5461	1,0000	0,990
Ambientefam	0,7261	1,0000	0,994
Educacionrec	0,0111	0,9950	0,974
Tegustacol	0,0111	0,9950	0,974
Haspenfu	0,0000	0,9950	0,973
Vicbullyng	0,0000	0,9950	0,973
Nivelespa	0,8825	0,9950	0,993
Cartracasa	0,0000	0,9560	0,935
Respetocodoc	0,2357	0,9950	0,979
Pactoconv	0,0000	0,9950	0,973

De los datos del Tabla 59, vemos en el Tabla 60 que aparecen organizados los valores de similitud (discriminante) del kernel para 2 grupos; aquí vemos que para cada varianza estimada hay 5 variables que tienen la mayor discriminancia; encontramos que las variables que muestran mayor poder discriminante para 2 grupos son **Perdidas**, **Entregatareas** y **Cartracasa**; además, pueden ser aceptada en menor ponderación para 2 grupos, **Tiempodedif**, ya que está incluida en

mayor proporción en diferentes estimaciones. Hay variables que están incluidas en buena proporción en el análisis, pero no son consideradas ya que se cree que no inciden en el actual estudio para 2 grupos.

Tabla 60. Valores kernel de la media y la desviación para variables discriminantes según la varianza. 2 grupos.

varianza 0.01			varianza 0.05		
Variables	Desviación	Media	Variables	Desviación	Media
Entregatareas	0,771	0,974	Entregatareas	0,933	0,988
Tiempodedif	0,900	0,987	Tiempodedif	0,972	0,997
Respetocodoc	0,949	0,979	Pactoconv	0,976	0,979
Pactoconv	0,967	0,973	Perdidas	0,976	0,969
Perdidas	0,969	0,935	Cartracasa	0,977	0,969

varianza 0.10			varianza 0.50		
Variables	Desviación	Media	Variables	Desviación	Media
Entregatareas	0,959	0,993	Haspenfu	0,985	0,994
Pactoconv	0,977	0,983	Pactoconv	0,987	0,995
Haspenfu	0,978	0,982	Entregatareas	0,989	0,998
Cartracasa	0,978	0,974	Perdidas	0,990	0,982
Perdidas	0,978	0,974	Cartracasa	0,990	0,982

varianza 0.70			varianza 1.00		
Variables	Desviación	Media	Variables	Desviación	Media
Haspenfu	0,988	0,995	Haspenfu	0,990	0,997
Pactoconv	0,990	0,996	Pactoconv	0,992	0,997
Perdidas	0,992	0,984	Perdidas	0,994	0,987
Cartracasa	0,992	0,985	Entregatareas	0,994	0,999
Entregatareas	0,992	0,999	Cartracasa	0,994	0,988

3.2.3. Discriminación de las variables por análisis discriminante y comparativo de grupos.

En la Tabla 61 que se presenta a continuación están agrupadas las 3 pruebas de igualdad de medias para 4, 3 y 2 grupos realizados en el análisis discriminante, de tal manera que se pueda hacer un paralelo del valor F para identificar las variables que entrarían al análisis en cada prueba y así compararlo en lo hecho con el método kernel. Recordemos que una variable puede hacer parte de la función discriminante si su valor de entrada F es 3,84 y expulsada si su valor de salida F es menos que 2,71 (AD).

Tabla 61. Comparativo de la Prueba de igualdad de medias para 4, 3 y 2 grupos.

Prueba de igualdad de medias de grupos									
	Prueba para 4 grupos			**Prueba para 3 grupos**			**Prueba para 2 grupos**		
Variables	Lambda de Wilks	F	Sig.	Lambda de Wilks	F	Sig.	Lambda de Wilks	F	Sig.
Perdidas	0,815	13,634	0,000	0,826	19,002	0,000	0,937	12,172	0,001
Tiempodedif	0,959	2,557	0,057	0,951	4,680	0,010	1,000	0,046	0,831
Entregatareas	0,936	4,084	0,008	0,961	3,651	0,028	0,998	0,362	0,548
Padrespend	0,968	1,954	0,123	0,959	3,905	0,022	0,971	5,348	0,022
Vivepadres	0,997	0,187	0,905	0,998	0,183	0,833	0,999	0,103	0,749
Ambientefam	0,996	0,211	0,889	0,998	0,221	0,802	1,000	0,020	0,888
Educacionrec	0,996	0,219	0,883	0,998	0,160	0,852	0,998	0,321	0,572
Tegustacol	0,993	0,397	0,756	0,998	0,153	0,858	0,998	0,284	0,595
Haspenfu	0,986	0,856	0,465	0,993	0,605	0,547	0,994	1,179	0,279
Vicbullyng	0,988	0,722	0,540	0,996	0,321	0,726	0,997	0,577	0,448
Nivelespa	0,997	0,198	0,897	0,999	0,131	0,877	1,000	0,033	0,856
Cartracasa	0,977	1,414	0,240	0,975	2,367	0,097	0,978	4,028	0,046
Respetocodoc	0,980	1,244	0,295	0,984	1,440	0,240	1,000	0,083	0,773
Pactoconv	0,988	0,702	0,552	0,992	0,721	0,488	0,995	0,900	0,344

Fuente: Diseño propio mejorado. Datos SPSS

Como vemos en la Tabla 61, para 4 grupos efectivamente solo entrarían dos variables a la función, **Perdidas y Entregatareas**. Para 3 grupos efectivamente entrarían 4 variables a la función, **Perdidas**, **Tiempodedif**, **Entregatareas** y **Padrespend**. Para 2 grupos efectivamente entrarían 3 variables a la función, **Perdidas**, **Padrespend** y **Cartracasa**.

Aunque el estadístico F es un poco riguroso con los valores de entrada y salida, se puede analizar la tabla completa y vemos que hay variables como **Entregatareas** y **Tiempodedif** que presentan buen valor promedio F y además, pueden tener buena influencia en el rendimiento académico, por lo cual deberían considerarse como ensayo en el análisis y función final.

Al analizar la escogencia de variables en el análisis discriminante y en el método kernel se identifica cierta similitud ya que se coincide en unas variables y hay otras que, aunque tienen una baja ponderación en últimas podrían ser incluidas por la influencia que podrían causar a la función; o sea que en este aspecto los dos métodos tienen parecido.

En el análisis final con dos grupos vemos que el análisis discriminante deja para la función las variables **Perdidas, Padrespend** y **Cartracasa.** El método kernel muestra el poder discriminante en las variables **Perdidas, Entregatareas** y **Cartracasa** pero las variables que son diferencia en una y en otra como **Padrespend** y **Entregatareas** han sido recomendadas por los valores arrojados y la influencia que pueden generar en este estudio.

3.2.4. Comparativo matriz de confusión (Resultados de la clasificación)

1. **Para el análisis discriminante tenemos que**:

Para 4 grupos se tiene el 60,30% de casos clasificados correctamente.

Tabla 62. Resultados de la clasificación AD, 4 grupos.

	Bajo	Basico	Alto	Superior	Total
Bajo	100%	0,00%	0,00%	0,00%	100%
Basico	9,00%	57,70%	18,60%	14,70%	100%
Alto	0,00%	10%	70%	20%	100%
Superior	0,00%	0,00%	25%	75%	100%
60,30% de casos agrupados o clasificados correctamente.					

Para 3 grupos se tiene el 68,50% de casos clasificados correctamente.

Tabla 63. Resultados de la clasificación AD, 3 grupos.

	Bajo	Medio	Alto	Total
Bajo	100%	0,00%	0,00%	100%
Medio	8,90%	64,60%	26,60%	100%
Alto	0,00%	9,10%	90,90%	100%
68,50% de casos agrupados o clasificados correctamente.				

Para 2 grupos se tiene el 91,30% de casos clasificados correctamente.

Tabla 64. Resultados de la clasificación AD, 2 grupos.

	Bajo	Aprueba	Total
Bajo	100%	0.0%	100%
Aprueba	8,90%	91,10%	100%
91,30% de casos clasificados correctamente.			

2. En el método kernel tenemos:

La Tabla 65, muestra una matriz donde se indica en paralelo los resultados de la clasificación realizados con la función kernel para diferentes valores de varianza con el fin de determinar cuál puede ser la varianza más eficiente para la clasificación en combinación con los resultados de la media. Es de aclarar que estas varianzas se tomaron en cuenta para diferentes ensayos.

Tabla 65. Resultados de la clasificación método kernel. 4, 3 y 2 grupos.

	Varianza									
Grupo	De cada Grupo	De la muestra	VM. De la muestra y Gru.	0,01	0,02	0,03	0,04	0,05	0,70	1,00
Bajo	75,00%	0,00%	100,00%	25,00%	75,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Básico	67,30%	66,30%	92,30%	14,10%	19,20%	25,00%	34,60%	34,60%	92,30%	100,00%
Alto	70,00%	0,00%	100,00%	35,00%	45,00%	70,00%	70,00%	80,00%	100,00%	100,00%
Superior	75,00%	0,00%	100,00%	25,00%	75,00%	75,00%	75,00%	75,00%	100,00%	100,00%
Total	67,90%	66,30%	93,50%	16,80%	24,50%	32,60%	40,80%	41,80%	93,50%	100,00%
Bajo	75,00%	0,00%	100,00%	25,00%	75,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Medio	66,50%	66,30%	91,10%	13,90%	19,00%	19,00%	24,70%	34,20%	91,10%	97,50%
Alto	81,80%	0,00%	95,50%	18,20%	18,20%	36,40%	36,40%	36,40%	95,50%	100,00%
Total	68,50%	66,30%	91,80%	14,70%	20,10%	22,80%	27,70%	35,90%	91,80%	97,80%
Bajo	75,00%	0,00%	100,00%	25,00%	75,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Aprueba	67,80%	67,80%	72,80%	12,80%	16,70%	16,70%	21,10%	25,00%	81,10%	87,20%
Total	67,90%	66,30%	73,40%	13,00%	17,90%	18,50%	22,80%	26,60%	81,50%	87,50%

En el Total de la Tabla 65, se observa el resultado de buena clasificación o clasificados correctamente mostrando que en los tres casos (4 grupos, 3 grupos y 2 grupos) las varianzas actúan arrojando datos proporcionalmente diferenciados para cada caso de grupos.

Para el caso de 4 grupos, en todas las varianzas se ve mejor eficiencia de clasificación y en su orden le sigue el de 3 grupos y después el de 2 grupos.

Analizando los valores estimados de varianzas, la mejor eficiencia de clasificación se ve en 0.50, 0.70 y 1.00, para los 3 casos, aunque se considera no tener tanto en cuenta la varianza 1.00 pues para este caso puede dar una clasificación errónea. Por último, para el caso de 2 grupos (Tabla 66) se hace uso de la probabilidad por tamaño de grupo con la función kernel y bajo este criterio se hace la clasificación utilizando diferentes varianzas.

Tabla 66. Clasificados correctamente de acuerdo a las varianzas por tamaño de grupo.

Clasificados correctamente de acuerdo a las varianzas			
Varianza	Porcentaje de buena clasificación	Varianza	Porcentaje de buena clasificación
0,01	22,83%	0,20	50,00%
0,02	26,63%	0,30	51,09%
0,03	34,24%	0,40	51,09%
0,04	34,24%	0,50	51,09%
0,05	38,04%	0,70	58,70%
0,10	42,93%	1,00	58,70%

Como se ve en la tabla los datos no son muy alentadores, ya que el porcentaje de buena clasificación va hasta 58,70% y si se sigue dando valores a la varianza en la modelación puede llegar hasta aproximadamente 61,0% de buena clasificación para este estudio. Aunque no se cumple el objetivo de clasificación por ser menor del 80%, resulta ser un buen ejercicio que podría ser aplicado en otros estudios, ya que nos permite identificar si es conveniente ponderar el centroide de acuerdo al tamaño de los grupos para hacer la clasificación, o si lo conveniente es no hacerlo.

Tabla 67. Resumen de Clasificados correctamente AD y Kernel

	Grupos	4	3	2	2 Ponderados
AD	**Varianza**	60,3%	68,5%	91,3%	
KERNEL	De cada Grupo	67,9%	68,5%	67,9%	
	De la muestra	66,3%	66,3%	66,3%	
	VM. De la muestra y Gru.	93,5%	91,8%	73,4%	51,09%
	0,010	16,8%	14,7%	13,0%	22,83%
	0,020	24,5%	20,1%	17,9%	26,63%
	0,030	32,6%	22,8%	18,5%	34,24%
	0,040	40,8%	27,7%	22,8%	34,24%
	0,050	41,8%	35,9%	26,6%	38,04%
	0,700	93,5%	91,8%	81,5%	58,70%
	1,000	100,0%	97,8%	87,5%	58,70%

En la Tabla 67 se hace un comparativo general, en donde se pueden observar en paralelo el porcentaje de individuos correctamente clasificados con el método kernel y el análisis

discriminante. Aquí se puede ver que en la última columna se hizo una ponderación de los grupos en torno a un centroide y la clasificación fue entre moderada y baja; por otro lado, el análisis discriminante solo muestra buena clasificación para 2 grupos y para más de 2 grupos es moderado; el método kernel presenta buena clasificación en varios valores de varianza (0.5, 0.7 y 1,0 respectivamente).

4. CONCLUSIONES

Se hace inicialmente el análisis discriminante para el rendimiento académico con todas sus descripciones para el estudio, y en todos los casos de grupos que se presentan, de tal manera que permita identificar toda su estructura, datos, los estadísticos, las funciones y los rendimientos de clasificación que le dan la validez.

Para la realización de este estudio originalmente se tomaron 15 variables y de estas una variable grupo o sea 14 variables predictoras. Indicadas como: Rendimiento académico, responsabilidad, acompañamiento, motivación, expectativa, convivencia, comunicación y cumplimiento de normas; aplicado a 184 estudiantes de 5 cursos de grado octavo.

Se realizaron 4 casos de clasificación: el primero para 4 grupos y como solo se llegó al 60,30% de correcta clasificación se procedió al siguiente caso; el segundo caso para 3 grupos llegando al 68,50% de correcta clasificación, pero se considera que aun el porcentaje es bajo, por tal motivo se procede a realizar el tercer caso para 2 grupos llegando al 91,30% de casos clasificados correctamente, esto indica que es una muy buena clasificación. Para los tres casos se realizó toda la estructura discriminante con el fin de conocer el comportamiento de los estadísticos aplicados, la clasificación de individuos en cada caso y las variables que predominan como discriminantes.

Realizado el proceso anterior se conoció que a menor cantidad de grupos y de variables el análisis discriminante se hace más eficiente. Ya conociendo los resultados para dos grupos se procede a realizar el cuarto caso, donde se utilizan dos grupos y las 3 variables de mayor poder discriminante según los datos obtenidos. Los grupos se forman de acuerdo a unos intervalos definidos por la necesidad del estudio a realizar.

Se pudo identificar en el análisis discriminante el estadístico F como elemento primordial para selección de variables y Lambda de Wilks como el estadístico que permite identificar el poder discriminante de una variable.

Los dos grupos que se clasificaron al final del proceso son Bajo y Aprueba con 86,40% de casos clasificados correctamente, lo cual es una muy buena clasificación; además, las variables que al final presentan el mayor poder discriminante son: **Perdidas**, **Padrespend** y **Cartracasa.**

Se halló e identificó los coeficientes de la función canónica discriminante estandarizada con los cuales se construye la función discriminante, con ayuda de las variables que presentan el mayor poder discriminante; además, con los coeficientes de la función discriminante canónica no tipificada, y los valores de las medias de las variables se puede llegar a calcular los centroides de los grupos, comparando y constatando el resultado hallado por el programa SPSS.

Con la probabilidad posteriori a partir de las puntuaciones discriminantes para 2 grupos se puede clasificar un individuo en el grupo 1 o 2 según el resultado de su respectiva función lineal.

Este método del AD es muy interesante porque permite clasificar los individuos en grupos por puntuaciones específicas de acuerdo a unas variables con poder discriminante y mediante el uso de unos estadísticos eficientes.

Para aplicar el método kernel como discriminante y clasificación de individuos se utiliza la fórmula del kernel Gaussiano en este caso; además, se hace uso de la media de los grupos, la varianza estimada de los grupos y la varianza de ensayo (0.01 a 1.00), con la variable promedio (variable grupo). Como el kernel genera medidas de similaridad de 0 a 1, se toma como base de clasificación el valor 0.60 (moderadamente significativa) para el grupo del cual se esté tratando; de esta forma se van incluyendo los individuos a los grupos, haciendo el conteo con el programa Excel. Este proceso se hace para 4, 3 y 2 grupos, para poder compararlo con lo hecho en el análisis discriminante. Entre los casos de grupos, se notó mayor eficiencia de clasificación para 4 grupos, y entre las varianzas utilizada la mayor eficiencia estuvo en la varianza 0.50 y 0.70 aproximadamente para 4 grupos (93.50%, 93.50%), 3 grupos (91.80%, 91.80%) y 2 grupos (73.40%, 81.50%).

Para identificar las variables de mayor poder discriminante se utiliza comparativamente, la media y la desviación estándar para cada grupo y variable, haciendo relación con los totales de cada variable; para esto se usa la fórmula del kernel Gaussiano. Solo se pensaba hacer con la desviación estándar, pero también se realiza con la media y se logró notar que resultados son casi iguales. Los valores cercanos a 0 (cero) indican buen poder discriminante y cercanos a 1

(uno) baja discriminación y si es 1 (uno) la discriminación es nula por que la similaridad es perfecta.

Para 4 grupos las variables más discriminantes son **Perdidas, Entregatareas y Tiempodedif;** para 3 grupos las variables más discriminantes son **Perdidas, Tiempodedif, Entregatareas y Padrespend;** para 2 grupos las variables más discriminantes son **Perdidas, Padrespend, Cartracasa** y **Entregatareas**.

De acuerdo a lo anterior, se logra notar que, para el análisis discriminante y para el método kernel las variables discriminantes halladas son muy similares para este estudio de rendimiento académico de los estudiantes de grado octavo.

En sentido general se ha notado que existen ciertas diferencias en resultado entre el método kernel y el análisis discriminante; diferencias en porcentaje de clasificación y algunas variables con poder discriminante, pero dependiendo la varianza que se asuma para aplicar en el método kernel, su resultado puede ser muy similar al análisis discriminante o podría mostrar mayor eficiencia.

Se puede decir que para clasificar variables e individuos conocidos y nuevos, existe mayor facilidad con el método kernel, pero se da la limitante para la aplicación de algunos estadísticos que pueden ayudar a facilitar y estructurar el proceso, para este caso, es necesario identificar valores semejantes o idénticos que conviertan al kernel en un estadístico similar para ese dato determinado.

El análisis discriminante mostró buena clasificación solo para 2 grupos y para más de 2 grupos fue moderado o medianamente eficiente.

Con el método kernel aplicado a dos grupos, cuando se ponderan el tamaño de los grupos en torno a un centroide, la clasificación de individuos es entre moderada y baja, ósea no es muy eficiente.

El método kernel puede presentar buena clasificación en varios valores de varianza estimada o de ensayo lo que nos puede generar la idea que, el método kernel es más eficiente que el análisis discriminante.

Las variables de mayor incidencia en la clasificación son Perdidas y Padrespend lo cual es un buen indicador de clasificación para este caso ya que el ganar las asignaturas y la dedicación de los padres son de vital importancia para evaluar el rendimiento académico; la variable

Cartracasa, es de gran influencia en el entorno o contexto que se esta estudiante (estrato 1, 2 y 3) ya que hay mucho estudiante que trabaja o realiza labores en casa, afectando de alguna forma su labor en el colegio; además, las variables Tiempodedif y Entregatareas resultan ser realmente importantes para el estudio en la evaluación del rendimiento.

Es importante aclarar que aunque hay variables que pueden incidir mucho se toma en cuenta el índice obtenido o el dato que arroja el estadístico quedando por fuera variables socioeconómicas o de convivencia muy importantes.

5. RECOMENDACIONES

Este estudio ha sido aplicado en la institución educativa técnica occidente de Tuluá Valle del Cauca, en grado octavo y la idea es que se haga extensivo a otros grados, toda la institución y a otras instituciones de la ciudad.

Haciendo un cambio de variables, este estudio puede ser aplicado a otras ramas del conocimiento, otros trabajos, otras investigaciones.

Para este estudio se hizo el comparativo del análisis discriminante y del método kernel aplicando una sola fórmula para todo el proceso; la sugerencia es que para próximos estudios se haga uso de las otras fórmulas del kernel y se comparen sus eficiencias.

Siendo el rendimiento académico un aspecto bien complejo la cual puede abarcar muchas variables, se recomienda asumir nuevos estudios como este, pero analizar otras variables y si es el caso confrontarlas con las variables de este estudio.

Asumir un estudio donde se midan las variables mediante otras escalas, lo cual podría dar resultados diferentes, y permitir otra forma de análisis.

6. BIBLIOGRAFÍA

[1] De la Fuente C. Laura (2012). (http://www.estadistica.net/Master - Econometría/Análisis_ Discriminante. pdf). Análisis discriminante. Pag 1-47.

[2] Carvajal, P., Mosquera, J., & Artamonova, I. (2009). Modelos de predicción del rendimiento académico en matemáticas I en la universidad Tecnológica de Pereira. Scientia et Technica, 258-263.

[3] Ariza B. Manuel (2006). Guía práctica de análisis de datos. Publisher, Junta de Andalucía, Consejería de Innovación, Ciencia y Empresa. Pag 93-100.

[4] Aldas M. Joaquín. (2005) El análisis discriminante. Universidad de Valencia.

[5] De la Fuente F. Santiago. (2011) Análisis discriminante. Facultad de Ciencias económicas y empresariales. Universidad Autónoma de Madrid. España.

[6] Edel N., Rubén (2003). El rendimiento académico: concepto, investigación y desarrollo. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, vol. 1, núm. 2. Red Iberoamericana de Investigación Sobre Cambio y Eficacia Escolar Madrid, España.

[7] Ramírez C., Juliana (2010) Regularización y Métodos Kernel para Algoritmos de Clasificación. Tesis de maestría. Facultad de Ciencias Exactas y Naturales. Universidad Nacional de Colombia. Manizales.

[8] Mendoza M., Adel A. y Herrera A., Roberto J. (2013). Propuesta para la predicción del rendimiento académico de los estudiantes de la Universidad del Atlántico,

basado en la aplicación del análisis discriminante. Artículo. Universidad del Atlántico. Barranquilla, Colombia.

[9] Manel M., Ramón, (2008) Introducción a los métodos Kernel. Universidad Autónoma y Carlos III de Madrid. Departamento de Teoría de la Señal y Comunicaciones. Madrid, España.

[10] Valle V., Carlos. (2009). Vector de Soporte y Métodos de Kernel. Universidad Técnica. Federico Santa María. Departamento de Informática. Santiago, Chile.

[11] Sánchez L. G., Osorio Germán A. y Suárez Julio F. (2008) Introducción a kernel ACP y otros métodos espectrales aplicados al aprendizaje no supervisado. Departamento de Matemáticas y Estadística, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Colombia. Manizales, Colombia.

[12] Peña, D. (2002), Análisis de datos multivariantes, McGraw-Hill. Pag 397- 428 y 449.

[13] Closas, A. H., Arriola E. A., Kuc, C. I., Amarilla, M.R., Jovanovich, E. C. (2013) Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. Enfoques XXV, 1 (otoño 2013): 65-92.

[14] Márquez M., Víctor A. (1989): "Apuntes sobre análisis muItivariante", Vol I. Mérida: Universidad de Los Andes. (Mimeo).

[15] Análisis discriminante (2001): El procedimiento Discriminante Capítulo 23

[16] Kernel Methods (2014). https://en.wikipedia.org/wiki/Kernel_method.

[17] Kovács K. (2008). Various Kernel Methods with Applications. Researcher Research Group on Applied Intelligence. University of Szeged. Doctoral School in Mathematics and Computer Science. Ph.D. Program in Informatics.

[18] Salvador Figueras, Manuel (2000). "Análisis Discriminante", [en línea] 5campus.com, Estadística <5campus.com/lección> [http://ciberconta.unizar.es/leccion/discri/100.HTM].

[19] Acuña Fernández, Edgar (2000). Notas de Análisis Discriminante. Departamento de Matemáticas, Universidad de Puerto Rico, Mayagüez, Puerto Rico.

[20] López D. Ana (2017). Fundamentos Matemáticos de los Métodos Kernel para Aprendizaje Supervisado. Facultad de matemáticas. Universidad de Sevilla. España. Pag. 40

[21] R. P. Wilson, A. G. Adams. (2013). Gaussian process kernels for pattern discovery and extrapolation. ICML, Pag. 1067 – 1075.

7. ANEXOS

ANEXO 1. Tabla de verificación de variables

CLASIFICACIÓN DE VARIABLES	VARIABLE	DESCRIPCIÓN	REPRESENTACIÓN	CATEGORIAS	RECODIFICACIÓN	CRITERIO
RENDIMIENTO	1. Nota promedio: Nivel de competencia	Cualidad producto de la calificación promedio obtenida en las notas de las áreas de todo el pensum académico en grado octavo. Se expresa en la variable con una escala de 1 a 10 puntos.	**Nivel** Variable Grupo	1. Bajo 2.Básico 3. Alto 4. Superior	1. (1,0 - 5,9) 2. (6,0 - 7,9) 3. (8,0 - 8,9) 4. (9,0 - 10,0)	La calificación mínima para el estudiante desde 1,0 por políticas de la institución.
	2**.** Asignaturas perdidas	Es una cualidad del estudiante que se obtiene producto de la cantidad de asignaturas perdidas o no haber perdido asignaturas.	**Pérdidas**	1. Deficiente 2.Regular 3. Bueno	1. (3 o más asignaturas perdidas) 2. (1 ó 2 asignaturas perdidas) 3. (No tiene asignaturas perdidas)	En **Bueno** el estudiante aprueba; en **Regular** se aprueba pero el estudiante debe de asumir un compromiso de recuperación; en **Deficiente** se reprueba pero el estudiante puede asumir un compromiso de recuperación para promoción anticipada al siguiente año.

RESPONSABILIDAD	3. Dedicas tiempo a estudiar por fuera de clase)	Cualidad del estudiante que refleja la dedicación por fuera de la clase	**Tiempodedif**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
	4. Entrega a tiempo sus tareas.	Cualidad del estudiante que refleja la responsabilidad para entregar tareas a tiempo	**Entregatareas**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
ACOMPAÑAMIENTO	5. Sus padres están al pendiente del desarrollo de sus actividades académicas.	En esta variable se identifica que tanto los padres están pendientes de las actividades académicas sus hijos y si apoyan en buena forma el proceso académico	**Padrespend**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
	6. Vives con tus padres.	Aquí se pretende saber si el estudiante vive con los padres o con otro tipo de acudiente, aspecto vital en el rendimiento académico.	**Vivepadres**	1. Un conocido. 2. Un familiar. 3. Uno de los padres. 4. Con los dos padres.	Los mismos valores de 1 a 4 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal

MOTIVACIÓN	7. Hay buen ambiente en su casa para desarrollar sus actividades académicas.	Esta variable muestra el nivel de convivencia que pueden estar pasando los estudiantes en sus casas	**Ambientefam**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
	8. Crees que la educación que estas recibiendo en el colegio es buena.	Aceptación de los procesos educativos que recibe el estudiante	**Educacionrec**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
ESPECTATIVA	9. Te gusta el colegio.	Esta variable muestra el gusto del estudiante por los elementos, normas y estructura de la institución educativa.	**Tegustacol**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
	10.Has pensando en tu futuro.	Aceptación, deseos y aspiraciones del estudiante a mejorar su calidad de	**Haspenfu**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su proyección de vida

		vida.		siempre. 5. Siempre.		
CONVIVENCIA	11. Es víctima de bullyng en el colegio.	En esta variable se identifica el trato a los estudiantes por parte de sus compañeros	**Vicbullyng**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
NIVEL FAMILIAR	12. Cuál es el nivel de escolaridad de tus padres o tu acudiente.	Esta variable mide el nivel de estudio de los padres o acudientes en el estudiante evaluado.	**Nivelespa**	1. No tiene estudio. 2. Primaria. 3. Bachillerato. 4. Pregrado. 5. Posgrado.	Los valores de 1 a 5 de acuerdo a la elección de la categoría presenciada en el estudiante.	Elección de la categoría de acuerdo a su criterio personal
	13 Le ponen excesiva carga de trabajo en su casa.	Esta variable permite determinar si el estudiante en su casa es sometido a largas horas de trabajo que no permitan su avance académico.	**Cartracasa**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal

CUMPLIMIENTO DE NORMAS	14. Tiene respeto por sus compañeros y docentes.	Variable que identifica el respeto del estudiante hacia la comunidad educativa.	**Respetocodoc**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal
	15. Reconoce el pacto de convivencia y lo cumple.	Variable que identifica si el estudiante reconoce los acuerdos estudiantiles dados en su matrícula.	**Pactoconv**	1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	Los mismos valores de 1 a 5 de acuerdo a la elección de la categoría	Elección de la categoría de acuerdo a su criterio personal

ANEXO 2. Formato de encuesta

	INSTITUCIÓN EDUCATIVA TECNICA OCCIDENTE DE TULUÁ Formato de encuesta sobre rendimiento académico	MUNICIPIO DE TULUÁ

Nombre completo del estudiante:
En qué grado está: _____ En qué Curso está: _____
Cuantos años tiene: _____ Sexo (M o F): _____

Marcar con una X donde corresponda

RENDIMIENTO	
1. Nota promedio: Nivel de competencia (Se saca del registro académico del colegio). 1. Superior_____(9,0 - 10,0) 2. Alto_____ (8,0 - 8,9) 3. Básico_____ (6,0 - 7,9) 4. Bajo_____ (1,0 - 5,9)	2. Asignaturas perdidas (Se saca del registro académico del colegio). Nota: Si perdió 3 o más asignaturas es Deficiente, si perdió 1 o 2 asignaturas es Regular, Si no ha perdido asignaturas es Bueno. 1. Deficiente (____) 2. Regular (_____) 3. Bueno (_____)
RESPONSABILIDAD	
3. Dedicas tiempo a estudiar por fuera de clase. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )	4. Entrega a tiempo sus tareas. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )
ACOMPAÑAMIENTO	
5. Sus padres están al pendiente del desarrollo de sus actividades académicas. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( )	6. Vives con tus padres. 1. Un conocido. ( ) 2. Un familiar. ( ) 3. Uno de los padres. ( )

4. Casi siempre. ( ) 5. Siempre. ( )	4. Con los dos padres. ( )
MOTIVACIÓN	
7. Hay buen ambiente en su casa para desarrollar sus actividades académicas. 1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.	8. Crees que la educación que estas recibiendo en el colegio es buena. 1. Nunca. 2. Muy pocas veces. 3. Algunas veces. 4. Casi siempre. 5. Siempre.
ESPECTATIVA	
9. Te gusta el colegio. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )	10. Has pensando en tu futuro. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )
CONVIVENCIA	
11. Es víctima de bullyng en el colegio. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )	
NIVEL FAMILIAR	
12. Cuál es el nivel de escolaridad de tus padres o tu acudiente. 1. No tiene estudio. ( ) 2. Primaria. ( ) 3. Bachillerato. ( ) 4. Pregrado. ( )	13. Le ponen excesiva carga de trabajo en su casa. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( )

5. Posgrado. ( )	5. Siempre. ( )
CUMPLIMIENTO DE NORMAS	
14. Tiene respeto por sus compañeros y docentes. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )	15. Reconoce el pacto de convivencia y lo cumple. 1. Nunca. ( ) 2. Muy pocas veces. ( ) 3. Algunas veces. ( ) 4. Casi siempre. ( ) 5. Siempre. ( )

ANEXO 3. Ejemplo para tabla de cálculo del kernel (programa Excel)

			Varianza y Media de Grupos					Varianza y Media del total de la muestra			
N°	PROMEDIO		N°	PROMEDIO	Exp	k(x,y)		N°	PROMEDIO	Exp	k(x,y)
1	6,40		102	5,60	-0,6806	0,50633562		102	5,60	-2,0052	0,13462763
2	6,20		42	5,70	-0,1250	0,88249690		42	5,70	-1,7377	0,17592256
3	8,10		106	5,90	-0,3472	0,70664828		106	5,90	-1,2601	0,28362891
4	7,80		132	5,90	-0,3472	0,70664828		132	5,90	-1,2601	0,28362891
5	7,80		96	6,00	-1,6286	0,19620850		96	6,00	-1,0500	0,34993883
6	7,40		13	6,10	-1,2757	0,27924678		13	6,10	-0,8591	0,42356350
7	6,40		30	6,10	-1,2757	0,27924678		30	6,10	-0,8591	0,42356350
8	6,50		93	6,10	-1,2757	0,27924678		93	6,10	-0,8591	0,42356350
9	6,60		97	6,10	-1,2757	0,27924678		97	6,10	-0,8591	0,42356350
10	7,80		108	6,10	-1,2757	0,27924678		108	6,10	-0,8591	0,42356350
11	6,80		2	6,20	-0,9658	0,38068354		2	6,20	-0,6873	0,50295565
12	7,10		83	6,20	-0,9658	0,38068354		83	6,20	-0,6873	0,50295565
13	6,10		92	6,20	-0,9658	0,38068354		92	6,20	-0,6873	0,50295565
14	9,40		111	6,20	-0,9658	0,38068354		111	6,20	-0,6873	0,50295565
15	6,90		23	6,30	-0,6990	0,49710209		23	6,30	-0,5346	0,58590286
16	8,10		26	6,30	-0,6990	0,49710209		26	6,30	-0,5346	0,58590286
17	6,90		27	6,30	-0,6990	0,49710209		27	6,30	-0,5346	0,58590286
18	7,80		46	6,30	-0,6990	0,49710209		46	6,30	-0,5346	0,58590286
19	8,50		51	6,30	-0,6990	0,49710209		51	6,30	-0,5346	0,58590286
20	9,00		74	6,30	-0,6990	0,49710209		74	6,30	-0,5346	0,58590286
21	9,00		76	6,30	-0,6990	0,49710209		76	6,30	-0,5346	0,58590286

ANEXO 4. Ejemplo de tablas para clasificación de individuos con datos del kernel para 4 grupos (programa Excel)

Grupo	Intervalo		Conteo	PROM. GRUPO	VARIANZA
Bajo	1,0	5,9	4	5,7750	0,0225
Básico	6,0	7,9	156	6,8699	0,2323
Alto	8,0	8,9	20	8,2700	0,0443
Superior	9,0	10,0	4	9,1250	0,0358
Total			184	7,0473	0,5223

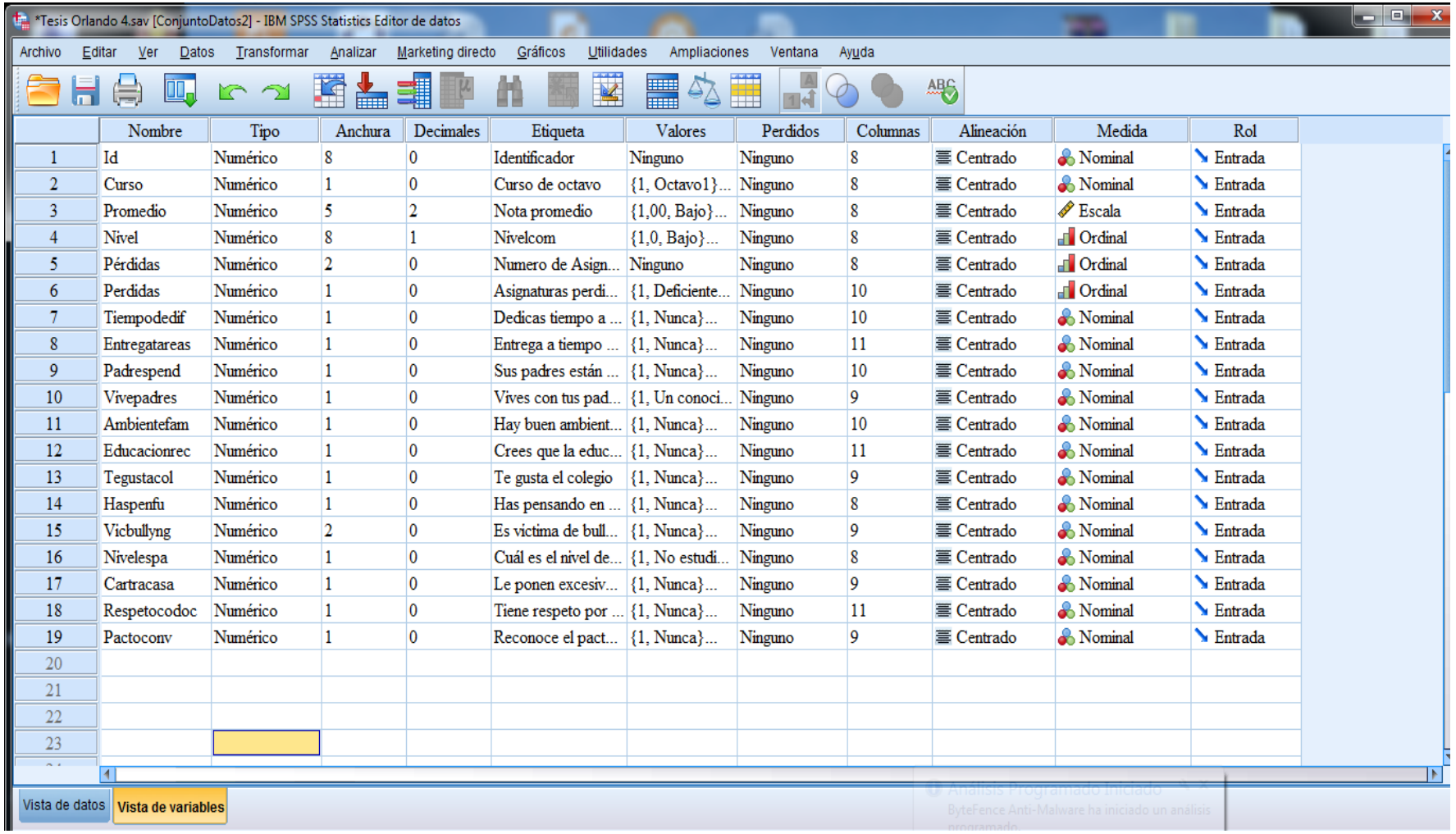
	Varianza									
Grupo	Grupo	Total Mue.	VM y Gru	0,01	0,02	0,03	0,04	0,05	0,7	1,0
Bajo	3	0	4	1	3	4	4	4	4	4
Básico	105	122	144	22	30	39	54	54	144	156
Alto	14	0	20	7	9	14	14	16	20	20
Superior	3	0	4	1	3	3	3	3	4	4
Total	125	122	172	31	45	60	75	77	172	184
Bajo	75,0%	0,0%	100,0%	25,0%	75,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Básico	67,3%	66,3%	92,3%	14,1%	19,2%	25,0%	34,6%	34,6%	92,3%	100,0%
Alto	70,0%	0,0%	100,0%	35,0%	45,0%	70,0%	70,0%	80,0%	100,0%	100,0%
Superior	75,0%	0,0%	100,0%	25,0%	75,0%	75,0%	75,0%	75,0%	100,0%	100,0%
Total	67,9%	66,3%	93,5%	16,8%	24,5%	32,6%	40,8%	41,8%	93,5%	100,0%

ANEXO 5. Ejemplo de tablas para clasificación de individuos del grupo real al grupo pronóstico (programa SPSS)

Estadísticas por casos

				Grupo superior				Segundo grupo superior			Puntuaciones
	Número del caso	Grupo real	Grupo pronosticado	P(D>d | G=g) p	gl	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado para centroide	Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado para centroide	Función 1
Original	1	2	2	,775	1	,972	,082	1	,028	7,152	,222
	2	2	2	,250	1	,727	1,321	1	,273	3,281	1,085
	3	2	2	,240	1	1,000	1,382	1	,000	17,107	-1,240
	4	2	2	,598	1	,944	,278	1	,056	5,919	,463
	5	2	2	,165	1	1,000	1,930	1	,000	18,919	-1,453
	6	2	2	,456	1	,999	,555	1	,001	13,733	-,810
	7	2	1**	,240	1	,712	1,381	2	,288	3,187	1,721
	8	2	2	,471	1	,999	,520	1	,001	13,553	-,785
	9	2	2	,840	1	,978	,041	1	,022	7,609	,138
	10	2	2	,616	1	,948	,252	1	,052	6,043	,438
	11	2	2	,962	1	,989	,002	1	,011	9,052	-,112
	12	2	2	,917	1	,983	,011	1	,017	8,161	,039
	13	2	1**	,186	1	,615	1,747	2	,385	2,686	1,574
	14	2	2	,560	1	,934	,339	1	,066	5,655	,518

ANEXO 6. Vista de variables para análisis discriminante (programa SPSS)



*Tesis Orlando 4.sav [ConjuntoDatos2] - IBM SPSS Statistics Editor de datos

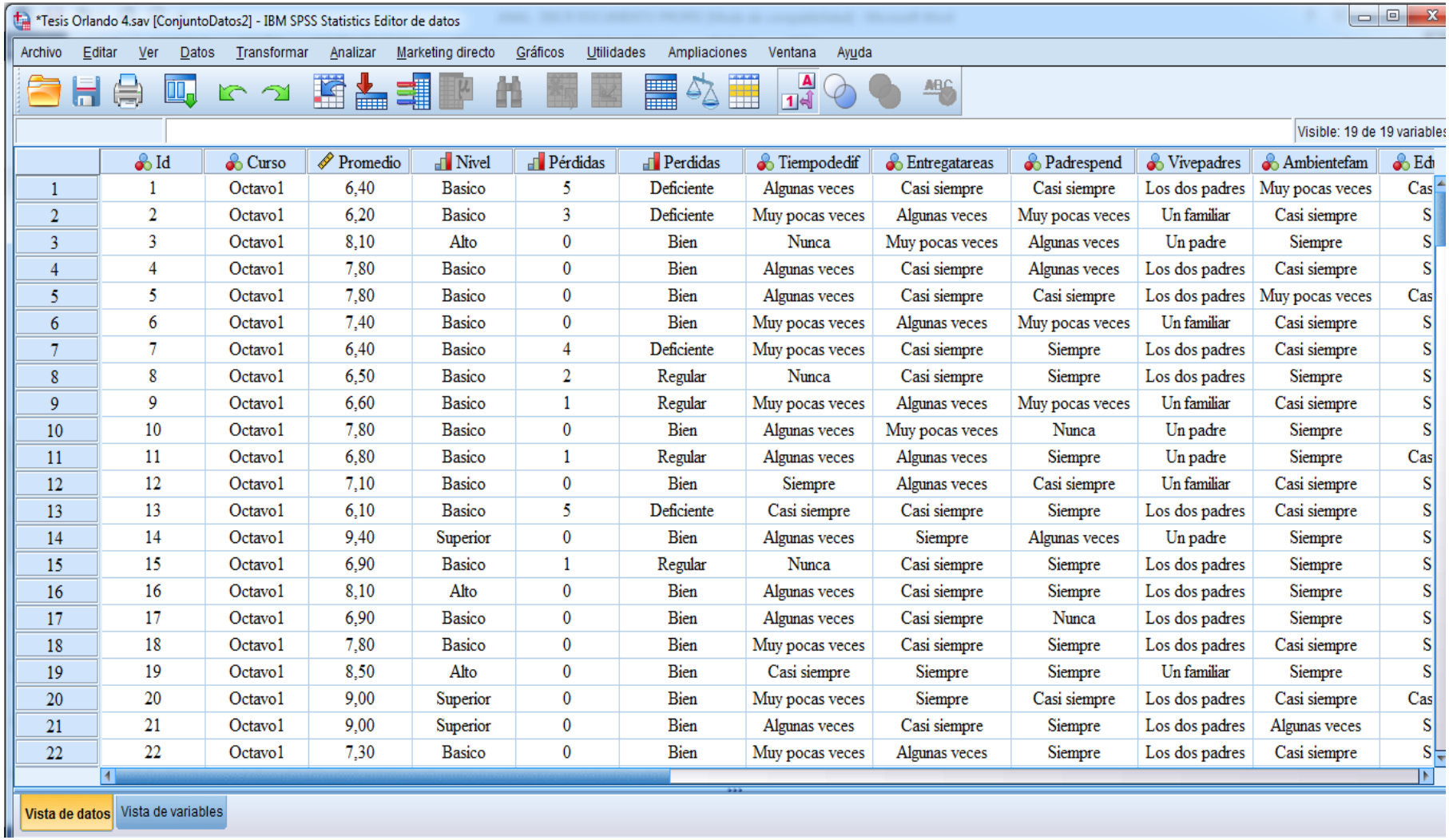
Archivo Editar Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ampliaciones Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Id	Numérico	8	0	Identificador	Ninguno	Ninguno	8	Centrado	Nominal	Entrada
2	Curso	Numérico	1	0	Curso de octavo	{1, Octavo1}...	Ninguno	8	Centrado	Nominal	Entrada
3	Promedio	Numérico	5	2	Nota promedio	{1,00, Bajo}...	Ninguno	8	Centrado	Escala	Entrada
4	Nivel	Numérico	8	1	Nivelcom	{1,0, Bajo}...	Ninguno	8	Centrado	Ordinal	Entrada
5	Pérdidas	Numérico	2	0	Numero de Asign...	Ninguno	Ninguno	8	Centrado	Ordinal	Entrada
6	Perdidas	Numérico	1	0	Asignaturas perdi...	{1, Deficiente...	Ninguno	10	Centrado	Ordinal	Entrada
7	Tiempodedif	Numérico	1	0	Dedicas tiempo a ...	{1, Nunca}...	Ninguno	10	Centrado	Nominal	Entrada
8	Entregatareas	Numérico	1	0	Entrega a tiempo ...	{1, Nunca}...	Ninguno	11	Centrado	Nominal	Entrada
9	Padrespend	Numérico	1	0	Sus padres están ...	{1, Nunca}...	Ninguno	10	Centrado	Nominal	Entrada
10	Vivepadres	Numérico	1	0	Vives con tus pad...	{1, Un conoci...	Ninguno	9	Centrado	Nominal	Entrada
11	Ambientefam	Numérico	1	0	Hay buen ambient...	{1, Nunca}...	Ninguno	10	Centrado	Nominal	Entrada
12	Educacionrec	Numérico	1	0	Crees que la educ...	{1, Nunca}...	Ninguno	11	Centrado	Nominal	Entrada
13	Tegustacol	Numérico	1	0	Te gusta el colegio	{1, Nunca}...	Ninguno	9	Centrado	Nominal	Entrada
14	Haspenfu	Numérico	1	0	Has pensando en ...	{1, Nunca}...	Ninguno	8	Centrado	Nominal	Entrada
15	Vicbullyng	Numérico	2	0	Es victima de bull...	{1, Nunca}...	Ninguno	9	Centrado	Nominal	Entrada
16	Nivelespa	Numérico	1	0	Cuál es el nivel de...	{1, No estudi...	Ninguno	8	Centrado	Nominal	Entrada
17	Cartracasa	Numérico	1	0	Le ponen excesiv...	{1, Nunca}...	Ninguno	9	Centrado	Nominal	Entrada
18	Respetocodoc	Numérico	1	0	Tiene respeto por ...	{1, Nunca}...	Ninguno	11	Centrado	Nominal	Entrada
19	Pactoconv	Numérico	1	0	Reconoce el pact...	{1, Nunca}...	Ninguno	9	Centrado	Nominal	Entrada
20											
21											
22											
23											

Vista de datos Vista de variables

Análisis Programado Iniciado
ByteFence Anti-Malware ha iniciado un análisis programado.

ANEXO 7. Ejemplo de Vista de datos para análisis discriminante (programa SPSS)



*Tesis Orlando 4.sav [ConjuntoDatos2] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ampliaciones Ventana Ayuda

Visible: 19 de 19 variables

	Id	Curso	Promedio	Nivel	Pérdidas	Perdidas	Tiempodedif	Entregatareas	Padrespend	Vivepadres	Ambientefam	Edı
1	1	Octavo1	6,40	Basico	5	Deficiente	Algunas veces	Casi siempre	Casi siempre	Los dos padres	Muy pocas veces	Cas
2	2	Octavo1	6,20	Basico	3	Deficiente	Muy pocas veces	Algunas veces	Muy pocas veces	Un familiar	Casi siempre	S
3	3	Octavo1	8,10	Alto	0	Bien	Nunca	Muy pocas veces	Algunas veces	Un padre	Siempre	S
4	4	Octavo1	7,80	Basico	0	Bien	Algunas veces	Casi siempre	Algunas veces	Los dos padres	Casi siempre	S
5	5	Octavo1	7,80	Basico	0	Bien	Algunas veces	Casi siempre	Casi siempre	Los dos padres	Muy pocas veces	Cas
6	6	Octavo1	7,40	Basico	0	Bien	Muy pocas veces	Algunas veces	Muy pocas veces	Un familiar	Casi siempre	S
7	7	Octavo1	6,40	Basico	4	Deficiente	Muy pocas veces	Casi siempre	Siempre	Los dos padres	Casi siempre	S
8	8	Octavo1	6,50	Basico	2	Regular	Nunca	Casi siempre	Siempre	Los dos padres	Siempre	S
9	9	Octavo1	6,60	Basico	1	Regular	Muy pocas veces	Algunas veces	Muy pocas veces	Un familiar	Casi siempre	S
10	10	Octavo1	7,80	Basico	0	Bien	Algunas veces	Muy pocas veces	Nunca	Un padre	Siempre	S
11	11	Octavo1	6,80	Basico	1	Regular	Algunas veces	Algunas veces	Siempre	Un padre	Siempre	Cas
12	12	Octavo1	7,10	Basico	0	Bien	Siempre	Algunas veces	Casi siempre	Un familiar	Casi siempre	S
13	13	Octavo1	6,10	Basico	5	Deficiente	Casi siempre	Casi siempre	Siempre	Los dos padres	Casi siempre	S
14	14	Octavo1	9,40	Superior	0	Bien	Algunas veces	Siempre	Algunas veces	Un padre	Siempre	S
15	15	Octavo1	6,90	Basico	1	Regular	Nunca	Casi siempre	Siempre	Los dos padres	Siempre	S
16	16	Octavo1	8,10	Alto	0	Bien	Algunas veces	Casi siempre	Siempre	Los dos padres	Siempre	S
17	17	Octavo1	6,90	Basico	0	Bien	Algunas veces	Casi siempre	Nunca	Los dos padres	Siempre	S
18	18	Octavo1	7,80	Basico	0	Bien	Muy pocas veces	Casi siempre	Siempre	Los dos padres	Casi siempre	S
19	19	Octavo1	8,50	Alto	0	Bien	Casi siempre	Siempre	Siempre	Un familiar	Siempre	S
20	20	Octavo1	9,00	Superior	0	Bien	Muy pocas veces	Siempre	Casi siempre	Los dos padres	Casi siempre	Cas
21	21	Octavo1	9,00	Superior	0	Bien	Algunas veces	Casi siempre	Siempre	Los dos padres	Algunas veces	S
22	22	Octavo1	7,30	Basico	0	Bien	Muy pocas veces	Algunas veces	Siempre	Los dos padres	Casi siempre	S

Vista de datos Vista de variables